

SPEECH MONOGRAPHS

Published by THE SPEECH ASSOCIATION OF AMERICA

Volume XXV

November, 1958

No. 4

Word Changes Introduced <i>Ad Libitum</i> in Five Speeches by Franklin Delano Roosevelt	229
LAURA CROWELL	
An Experimental Investigation of Rhetorical Clarity	243
ROGER E. NEBERGALL	
A Possible Basis for the Association of Voice Characteristic and Personality Traits	255
EDITH B. MALLORY and VIRGINIA R. MILLER	
Personality and Discussion Behavior: A Study of Possible Relationships	261
THOMAS M. SCHEIDEL, LAURA CROWELL, and JOHN R. SHEPHERD	
Experimental Studies of Motivated Group Discussion	268
WILLIAM E. UTTERBACK and WALLACE C. FOTHERINGHAM	
Effects of Certain Techniques of Credibility Upon Audience Attitude	278
THOMAS S. LUDLUM	
The Relationship between the Ability to Speak Effectively and the Primary Mental Abilities, Verbal Comprehension and General Reasoning	285
JOE M. BALL	
Oral and Nasal Sound Pressure Levels of Vowels	291
RAYMOND SUMMERS	
Observed and Predicted Estimates of Reliability of Aspects of a Speech Articulation Rating Scale	296
NANCY MARCH, CARL H. WEAVER, SHEILA MORRISON, and JOHN W. BLACK	
INDEX TO VOLUME XXV	305

SUGGESTIONS FOR CONTRIBUTORS TO SPEECH MONOGRAPHS

A portion of *SPEECH MONOGRAPHS* is devoted to the publication of articles based on original research, representing the various areas of specialization and techniques of investigation included in the field of speech.

These reports should be given in enough detail to permit the reader to know how the author carried out his investigation and how he arrived at his conclusions. Sources of evidence, conditions of observations, methods of gathering data, should be cited or described. Reports of experimental investigations should follow the usual pattern: statement of purpose, procedures used in gathering data, analysis of data, conclusions. While the inclusion of the data is essential, it should be presented concisely. Tables should be on separate pages. If statistical methods are employed, they should be named, but they need not be described if they are standard, or are adequately explained in some source to which the author can refer.

Articles in *SPEECH MONOGRAPHS* vary considerably in length. Short ones will be considered by the Editorial Board when they represent investigations the nature or scope of which permits a brief report without omitting essential details. Authors preparing reports of major studies involving extensive data and/or analysis, and anticipating that their material will exceed 8,000 words, should write to the Editor, inquiring about the availability of space and giving an estimate of the probable length of the material.

Contributors should have their manuscripts read by competent critics before submitting them for publication. When manuscripts are based on dissertations, that fact should be indicated, with the name of the director of the original research, in a footnote. In such cases, the director should review the manuscript. All copy should be proofread by at least one person besides the author before being submitted for publication. The number of words in the copy should be noted in the upper right-hand corner of the first page. Titles should be typed in full capitals, with the author's name in capitals, two spaces below. The author's affiliation, school or college (or home city, if he is not so affiliated), should appear, with normal capitalization, two spaces below his name. Footnotes should be typed on separate sheets following the last page of the manuscript. They should be double spaced, with triple spacing between items. Authors should retain a duplicate manuscript, and should send a carbon copy to the editor along with the original. The original copy should be on heavy bond paper.

All manuscripts must conform to the *MLA Style Sheet*, a copy of which can be secured from the Treasurer, Modern Language Association, 100 Washington Square East, New York 3, New York. Price: 25 cents.

Manuscripts will be read by the Editor and members of the Editorial Board before a final decision is made. Accepted manuscripts will be published in approximately the order of their acceptance. The deadline for submitting final revision of accepted manuscripts is two months before date of issue. Publication dates are March, June, August, and November. Contributors should keep the Editor informed of any changes in address.

Reprint order blanks are sent to authors; orders should be placed at once with the Arcraft Press, Columbia, Missouri. Reprints cannot be secured later.

Manuscripts should be sent to John W. Black, Editor of *SPEECH MONOGRAPHS*, Department of Speech, The Ohio State University, Columbus 10, Ohio.

SPEECH MONOGRAPHS

VOLUME XXV

NOVEMBER, 1958

No. 4

WORD CHANGES INTRODUCED *AD LIBITUM* IN FIVE SPEECHES BY FRANKLIN DELANO ROOSEVELT*

LAURA CROWELL
University of Washington

FRANKLIN ROOSEVELT worked diligently on his speeches. That he regularly provided the basic thoughts which he wanted incorporated in an address—indeed, often dictated a few pages of his own thinking as a starting-point—and then worked again and again with his speech assistants to bring the manuscript to the precise length, content and tone he desired is now widely understood. Both the anticipated immediate audience and the thousands of unseen radio listeners were considered in this exhilarating, exhausting process of readying the final draft; Roosevelt also often read the developing paragraphs aloud to test their oral quality.

In Roosevelt's addresses we have particularly apt material for a study of verbal changes made during the act of speaking. We are able to know what he planned to say, for the Reading Copies of his major addresses are available for study at the Hyde Park Library. We are able to know also what he actually said, for the electrical transcriptions of his speeches, which reveal his spontaneous alterations, are available from the National Archives.

One would not expect to discover any large number of misreadings because

* The electrical transcriptions were obtained through funds from the Agnes Anderson Fund, University of Washington.

this was a presentation of well-known material by a speaker who took pride in reading well. Furthermore, no one knew better than did Roosevelt that if any spontaneous alteration were to result in an error of fact, a shallowness of inference, or an indication of stigma or unintentional sarcasm, he would have put a weapon in the hands of ready and gleeful attackers of him and his policies and often, indeed, of his country as well. Another factor which might have held back his deliberate ad-libs was the urging of his speech assistants that he stick to the text which had been so painstakingly prepared.¹

Yet during the presentation of even major addresses Roosevelt did make changes. A careful examination of his delivery of five selected addresses to large immediate audiences and to the country by radio will reveal the nature and extent of his ad-libbing and provide a basis for speculation as to his reasons for departing from the words of the manuscripts.

What was the nature of the audience before whom these addresses were given,

¹ See James M. Burns, *Roosevelt: The Lion and the Fox* (New York: Harcourt, Brace, 1956), p. 421; Samuel I. Rosenman, *Working With Roosevelt* (New York: Harper & Bros., 1952), p. 486; Robert E. Sherwood, *Roosevelt and Hopkins* (New York: Harper & Bros., 1950), p. 218.

the throngs of listeners whose thunderous presence and response were a counterstimulation to the President? The millions beside their radios were not a part of this vigorous interaction; adaptation to their anticipated attitudes had been woven into the fabric and fashioning of the addresses by FDR and the speech assistants, but adaptation to the thousands before him was a continuing matter and, therefore, the focus of his efforts in spontaneous alteration.

Some 60,000 people jammed into the National League Baseball field in Pittsburgh for the October 1, 1936, campaign address; ardent Democrats, noisily celebrating their success of the previous November heard the Victory Dinner Address in Washington, D. C., on March 4, 1937; a huge partisan crowd thundered applause at his rhythmic jibe at "Martin, Barton and Fish" in the excitement-rouser at Madison Square Garden, October 28, 1940; the Houses of Congress, assembled for the annual presidential message to Congress, heard Roosevelt explain the Lend-Lease Plan and set forth the Four Freedoms on January 6, 1941; enthusiastic supporters, one thousand strong, in a crowded banquet hall in the Statler Hotel delighted in his masterful telling of the Fala incident in the Teamsters' Union Address in Washington, D. C. September 23, 1944.² These, then, were the situations: three campaign rallies, a political dinner meeting, an assembled Congress. It was in oral presentation before these groups and in adjustment to their responses that FDR made the ad-libs to be studied here.

By ad-libbing in these addresses we mean substitutions, additions, insertions, deletions, transpositions from the

² Rosenman, who worked for 17 years with FDR on his speeches, considered this address the best political campaign speech he ever made. *Op. cit.*, p. 248.

copy he had in hand. Taking rough account of the varied lengths of the speeches and using the Pittsburgh address with its 172 instances of spontaneous alteration as a basis, we find that the other two campaign addresses had nearly as many. On the other hand, the Victory Dinner address to a friendly audience but on the hazardous Supreme Court issue had only one-half as many, and the Four Freedoms address on a safer theme but to the more formal Congressional audience had only one-third as many.³ Thus, of these five addresses, the campaign speeches seemed to allow him greater freedom in ad-libbing than did the other occasions.

Detailed analysis of these ad-libs shows them to fall into two major classes. Roosevelt appears to have continued under the impact of the speaking situation itself the work which he and his speech assistants had pursued so assiduously in their labors on the drafts, the result being merely a further tailoring of the text. But in the oral presentation before the massed listeners and in the very act of communication other changes occurred, changes directly associated with this immediate speaker-audience relationship; these alterations form a second type of ad-libs.

FURTHER TAILORING AD-LIBS

In an earlier study of the Four Freedoms address it was discovered that the alterations made during the working of the seven drafts arranged themselves in five categories: *relationships*, *concreteness*, *force*, *motivation*, and *composition*.⁴ It seems logical that if we now

³ These comparisons indicate Roosevelt's amount of ad-libbing in only the most general way, of course, since they refer to the number of times he deviated from the text in any way and not to the nature or extent of the introduced or omitted material.

⁴ Laura Crowell, "The Building of the 'Four Freedoms' Address," *Speech Monographs*, XXII (November 1955), 266-283.

find ad-libs falling into these same categories we may conclude that FDR in the very act of presentation was truly continuing the work which had been halted by the necessity of binding the Reading Copy (now typed triple-space on thin, non-rustling sheets) in the limp black folder ready for use on the platform.

Relationships

Roosevelt's desire to make his statements inclusive, to avoid neglect of any element, continued within the actual presentation of the speeches: in declaring that his Administration had planned to protect "the small business, the small corporation, the small shop" from deflation, he made the impromptu addition, "and the small individual" (Pitts-13);⁵ in speaking of the faith of Americans in "political and economic democracy," he inserted "and social" (Vic-3); in asserting that without repeal of the embargo no help would have been afforded Britain, he expanded by adding "and a lot of other nations" (MSG-55). It seems that his drive to be inclusive remained vigorously at work while he was actually addressing the audience.

Despite FDR's liking for specificity, his relational, movement sense was perhaps even stronger; for example, in referring to an opponent's speech, he altered his text from "in June, 1938" to "a little over two years ago" (MSG-22). Furthermore, what was implicit through the sentence structure provided in the Reading Copy FDR sometimes made explicit in his oral presentation by adding

⁵ Throughout the manuscript reference to the speeches will be made as follows: *Pitts* for the Pittsburgh address; *Vic* for the Victory Dinner address; *MSG* for the Madison Square Garden speech; *FF* for the Four Freedoms address; *TU* for the Teamsters' Union address. The numbers will designate the paragraphs as presented in *The Public Papers and Addresses of Franklin D. Roosevelt*, compiled by Samuel I. Rosenman.

the relation-giving word spontaneously; he ad-libbed to express relationships in these sentences: "I rejected that advice because nature was in an angry mood" (Pitts-11);⁶ "efforts of this Administration to arm itself adequately for purposes of defense" (MSG-10); "when government finance or any other kind of finance is honest" (Pitts-2). He also verbalized relationships for which words were less clearly suggested by the Reading Copy; for instance, speaking of the cooperation which able men of other politics were giving the Administration, FDR then ad-libbed as follows: "And, the other way 'round, this Government is cooperating with them—one hundred per cent" (MSG-12).

These motivations to clarify relationships, so strongly felt by FDR and at work all the time during the preparation of the manuscript, seem to have continued their influence while he was in the act of speaking.

Concreteness

In spite of the care taken at every stage of preparation of the drafts to put ideas in specific terms, FDR nevertheless in a few instances added details while speaking. These details were sometimes geographical, as the addition of Louisville and Memphis to his statement about "men piling sandbags on the levees at Cairo" (Vic-54), and of the Connecticut River when speaking of the extension of flood control projects (Vic-51). Sometimes the details added were even numerical, as when the script referred to the amount of American money sent abroad in a single decade as "not merely eight billions but many

⁶ Where note is to be taken within quotations from the addresses of alterations made in the act of speaking, all changes except omissions will be italicized; omitted elements will be indicated by parentheses. For example, here the word *because* was ad-libbed.

more billions" and FDR added "about fourteen billions" (Pitts-28).

Again, the details added spelled out a purpose more specifically; for example, when speaking of "landing vessels" as one of the types of war equipment that American industry had produced, he inserted the explanation, "to get the troops ashore" (TU-22). He brought abstractions down to earth with concrete elements: after the phrases "to restore purchasing power, to raise values" he added "to put people back to work" (Pitts-19).

Thus, FDR continued to alter his speeches during their presentation in the direction of greater concreteness.

Force

Despite the concerted attempt by Roosevelt and his speech writers to develop the strongest verbal and syntactical instrument possible he yet found occasion in these addresses to add words to gain strength for the vocalization of the ideas. At one point he added the idea of desirability to that of capability when he inserted a vigorous clause: "Since we could not get the money by taxes and ought not to have got it by taxes in those years we borrowed it" (Pitts-23).

When the words in the text did not seem capable of carrying the vocal weight of the idea he was trying to express, he ad-libbed to provide himself a more suitable verbal vehicle. Sometimes the ad-libs merely asserted the strength of the idea: "The Republican leaders were definitely and beyond peradventure of doubt trying to block our efforts toward national defense" (MSG-19). At times words were added to permit the amplification of an idea, to provide for an expansiveness he was feeling as he spoke, as in these examples: "progress that's being made in the whole picture of our national defense" (MSG-

14); "production of all the types of machines and tools and steel and everything else that have made this nation the industrial leader of the world" (MSG-11). Elsewhere the addition of the time element was used to fulfill this need for amplification: "That determination of ours, extending over all these years, was proved. . ." (FF-5).

In another instance, the word "roar" even when preceded by a highly significant pause did not seem to offer phonetic scope for the size of his idea, and FDR held attention on its significance by pointing at it verbally: "This new roar—for that's the best term—is the best evidence in the world" (Vic-9). And once he spelled out a key word as well as repeated it to build force: "They were botched b-o-t-c-h-e-d they were botched by a Republican Administration" (TU-43). At one point he introduced the word "why" as an expletive, as though verbalizing thereby a scope for his idea not sufficiently allowed by the Reading Copy words: "'Don't leave the task of making the peace to those old men . . . why just turn it all over to us'" (TU-15).

FDR often supplied additional words at the outset of sentence or clause to serve a general leading-in, introductory purpose; the most commonly added word at the sentence opening was "and." In these examples the addition of the word "and" provided a malleable opening syllable and allowed the building up of force for the important word that followed: "And they are playing politics with the national security of America today" (MSG-30); "And tonight I'm going to talk to you about the boxscore of the government of the United States" (Pitts-1). But this introductory function was also often served by more repetition of the opening words of the sentence; these examples show Roosevelt winding up for more vigorous utterance: "The

Government *the Government* of this great Nation, solvent, sound in credit, is coming through . . ." (Pitts-42); "The first thing *the first thing* they did was to try to eliminate the battleships" (MSG-42).

At the beginning of clauses within sentences the word "that" was often added to play this introductory role, as in this example: "candidates . . . still think *that* you ought to be good boys and stay out of politics" (TU-24). Again the repetition of modifiers or prepositions in a series provided this opening element: "when *the* incomes and *the* values and *the* transactions of *the country* are on the down-grade" (Pitts-7); "it is the same record of timidity, *of* weakness and *of* shortsightedness" (MSG-64).

Often in presenting a series orally FDR reached a step farther back in the structure of the sentence to project the parallel elements more vigorously; for example, "You and I and everybody else saw millions out of work, *saw* the business concerns running in the red, *saw* the banks closing" (Pitts-5); and "It is necessary to mobilize resources, *to mobilize* minds and skills" (MSG-94).

Just as FDR ad-libbed to provide words to serve a building-up function, so also he added elements to round out the vocal emphasis which he had built up for ideas; these words seem added as a sort of retiring, subsiding action: "In fact, it looked so much like a shut-out *for the team* that you voted a change in management" (Pitts-3); "they keep this business up usually for three years and six months *in a row*" (TU-3); "And

if I have ought to say *about it*, it will continue" (Vic-7); "Modern science . . . knows how to produce as a by-product the blessing of cheaper electric power *for the Nation*" (Vic-50).

Not only did FDR add elements to put vocal emphasis where he, as he spoke, wished to put it, but he eliminated syllables that might have defeated his intended stress: "To balance our budget in 1933 or (19)34 or (19)35 would have been a crime against the American people" (Pitts-16). Furthermore, FDR at times enlarged an idea sufficiently by vocal intonation and pause alone, thus making unnecessary words which might otherwise have carried the emphasis: "The people (themselves) understood that in 1933" (Pitts-22); "Listen to this (statement) for instance: (I quote)" (MSG-21). Elsewhere he seemed to have spent his force before completing the Reading Copy sentence and thus omitted the tapering-off modification: "we move forward with God's help to the greatest epoch of free achievement by free men *that* the world has ever known (or imagined possible) (TU-56).

Thus, in the act of speaking, Roosevelt added or deleted elements as he strove for a verbal instrument capable of expressing the ideas with the strength and scope which he desired.

Motivation

During presentation of the addresses, Roosevelt heightened the affective power of his ideas by several forms of verbal change. Sometimes the increase came from the substitution of a word or phrase of stronger connotation:

from
so long as it continues to make modern democracy work

to
so long as it continues to make a modern democracy succeed . . . (Vic-6)

He invigorated a sentence, changing
from

Please God, may that spirit never prevail in our
land

He roughened a final admonition

from
foreign boondoggling, if you will

Sometimes he strengthened feeling by spelling out commonly desired goals, as when he followed the Reading Copy pronouncement that "there has been no entanglement" with the prediction and promise for the future, "and there will be no entanglement" (MSG-75).

FDR further introduced verbal asides which evoked or permitted overt expression of the listeners' feelings. On the one hand, he animated them by voicing a matter of common rejoicing as, for example, in the successful prosecution of the war: "And I may say that those Commanding Generals are making good in a big way" (TU-18). On the other, he aroused them by citing a matter of common derogation: shouts and a cry of "You tell 'em, Frankie" followed his punch-like line regarding taxes, "To pile (on) vast new taxes would get us nowhere because values were going down, *and that makes sense too*" (Pitts-9); he lifted the booing and laughter to further heights by repeating with great sarcasm of tone the element they had enjoyed, "Senator Taft—the runner-up *the runner-up*" (MSG-25); he followed a jibe which had brought applause and shouts with an ad-libbed element of great belligerence, "for I am not through yet" (Vic-45), and drew more laughing and whistling. The well-known Fala story was further tuned to the audience's wild appreciation by ad-libs throughout:

from

These Republican leaders have not been content with attacks upon me, or my wife or my sons—they now include my little dog, Fala.

to

Pray God that, having discovered it over here,
we won't forget it either (MSG-67).

to a savage

foreign boondoggling, if you like it (Pitts-28).

Often Roosevelt augmented the sense of immediate involvement by his spontaneous changes. Sometimes the ad-lib belabored an idea to bring it more dynamically to the consideration of the listeners, as in the matter of time: "But last fall—*this fall, 1940*, they had their chance" (MSG-49); "this Administration—*this one—today*—is responsible" (TU-32). Sometimes the ad-lib returned an idea discussed earlier to their attention sharply: "We must set targets and objectives for the future which will seem impossible—*like the airplanes*—to those who live in . . . the dead past" (TU-53). Elsewhere he made direct requests for their consideration of an idea: by ad-libbing complete sentences, "Remember, a number of years ago, there was a book Mein Kampf written by Hitler himself" (TU-31), and "Well, let's take some simple illustrations that come to mind" (TU-32); by inserting elements, as "look at their names" (TU-37), or rephrasing ideas more vigorously, as in the substitution of "Turn another page" for "one more example" (MSG-49). At times he asserted their real involvement by such inserted elements as these: "We knew *and you knew* in March, 1933, that it would cost money" (Pitts-21), and "Something had to be done, *we all know that*" (Pitts-11).

At many points FDR ad-libbed elements which served to formalize the message still further. At times this was

to

These Republican leaders have not been content with attacks on me, or on my wife, or on my sons—no, not content with that—they now

Unlike the members of my family, he resents this. Being a Scottie, as soon as he learned that the Republican fiction-writers had concocted a story that I had left him behind on an Aleutian Island and had sent a destroyer back to find him—at a cost to the taxpayers of two or three or twenty million dollars—his Scotch soul was furious. He has not been the same dog since.

include my little dog, Fala. Well, of course, I don't resent attacks, and my family don't resent attacks, but Fala does resent them. You know—you know, Fala is Scotch and being a Scottie, as soon as he learned that the Republican fiction-writers in Congress and out had concocted a story that I'd left him behind on an Aleutian Island and had sent a destroyer back to find him—at a cost to the taxpayers of two or three or eight or twenty million dollars—his Scotch soul was furious. He has not been the same dog since. (TU-38)

done by yielding the audience confidence, as in "oh, I can't understand it, I don't believe you can either" (TU-22), and in "my, it seems a long time ago" (TU-13). Elsewhere he brought the mechanics of speech-making forthrightly into view, as when he ad-libbed "To use the old example that's always good" (MSG-39), or called informally upon authorities, "You see, I'm talking by the book" (MSG-44), or used a prominent name familiarly, "And I think it was a kinsman of mine, about 30 years ago, who invented the word, 'weasel words'" (MSG-4). Often he obtained increased informality by ad-libbed uses of colloquial language: "Can you beat that?" (TU-23); "You know, they pop up all the time" (TU-34); "I've learned by now to take it on the chin" (MSG-79).

Roosevelt heightened the emotional impact of his ideas as he addressed his audience by substituting and adding elements which evoked popular feeling, augmented the sense of immediate involvement and increased further the close rapport between speaker and listeners.

Composition

Certain types of changes made during the presentation of the speeches may profitably be considered as merely compositional, although they also serve the general aim of vigorous communica-

tion.⁷ For example, the Reading Copies were consistently prepared without the informal pronoun-verb contraction; FDR almost as consistently altered these forms as he spoke: I am becomes *I'm*; it is, *it's*; we had, *we'd*; who have, *who've*. The informalizing effect of such contractions is well recognized; much less attention has been given to the fact that this practice allowed him to move quickly into the succeeding phrases and to be in step with their movement and nature; for example, the replacement of "We have all seen" accelerated the beginning of this sentence: "We've seen many marvelous stunts in the circus, but no performing elephant could turn a handspring without falling flat on the back" (TU-11).

Another alteration clearly related to ease of sentence-movement is the vocal substitution of *that* for *which*; this alteration occurred twenty-one times in

⁷ In a general sense all adlibs could be classified as compositional, since all are alterations of phraseology to enable the speaker to communicate ideas and attitudes more fully. Nevertheless, certain changes are unique to their context, arising out of the ideas to be communicated, and serving purposes of clarifying relationships, giving concreteness, providing force or adding motivation; others, however, though also serving these same purposes, take again and again the same forms and establish themselves as a frequently-found type of change. Therefore, the content-centered group has been already considered above under appropriate headings; the form-centered group will now be taken up as compositional changes.

the five speeches under consideration.⁸ Repeated listening of FDR's transcribed addresses gives one the impression that the phonetic components of the word *that* allow for much greater adjustment than those of the word *which*; *that* can be tucked in vocally under other words and thus allow the stress to fall on them, but *which* is not capable of such swift, easily-modified production.⁹ Clearly the superior phonetic flexibility of the one word is the reason for its substitution in sentences like this: "Compare the score-board *that* (which) you have in Pittsburgh now with the scoreboard *that* (which) you had when I stood here at second base in this field four years ago" (Pitts-36).

Another instance of FDR's substitution for the sake of vocal emphasis occurs in his interchange of the words *on* and *upon*. It appears that when he wished to place particular stress upon the substantive words of the phrase he minimized the relational words, substituting *on* for *upon*.¹⁰ In five cases in these addresses FDR substituted *on* orally for the *upon* shown in the Reading Copies; for example, "the crushing

⁸ Although FDR sometimes wrote word changes into the Reading Copy with his pen, at no point in these five addresses had he so struck out the *which* and written in a *that*; these were spontaneous oral changes.

⁹ Indeed, with [ɪ] relatively incapable of reduction to a more relaxed vowel in comparison with [æ], and the complexity of the closing consonant [tʃ] in comparison with the simplicity of [t], one can see that FDR would be able to serve his exquisite sense of timing more adequately with *that* than with *which*. An auditory analysis of Roosevelt's Address to the Congress on January 3, 1936, noted his "shortening of unemphasized material." Charles H. Voelker, "A Phonetic Study of Roosevelt," *The Quarterly Journal of Speech*, XXII (October 1936), 367.

¹⁰ H. W. Fowler has explained that if euphony would cause *upon* to have two unaccented syllables instead of being produced with a clear ō, then *on* is usually preferred. *A Dictionary of Modern English Usage* (Oxford University Press, 1950), p. 683.

load the debt will impose on (upon) your children and mine" (Pitts-37); "The happiness of future generations of Americans may well depend on (upon) how effective and how immediate we can make our aid felt" (FF-56); "a plan based on (upon) the wishes of the soldiers themselves" (TU-35). But he also varied the words in the opposite manner. In one case Roosevelt had written into the Reading Copy by his own hand a sentence closing with the words "an attack on us" (MSG-18); nevertheless, in speaking, he substituted *upon* for the written *on*, and used full stress on the word's final syllable. Clearly he made spontaneous alterations to allow himself the vocal emphasis he desired.

A frequent change related to sense rather than to sound involves the words *this* and *that* (also *these* and *those*). He shifted from *this* to *that* and from *these* to *those* to set aside a completed thought and move into a succeeding idea. For example, he refocused attention by such substitution: "Contrast *those* (these) expenditures and what we got for them with certain other expenditures" (Pitts-28). At other times a deprecatory force seemed to underlie the substitution as if removing the undesirable to a greater distance: "six billion dollars of increased assets to set off against *those* (these) liabilities" (Pitts-33). He changed also in the opposite direction, as, for example, when he drew an already mentioned element more clearly into focus: "Tonight I want to begin with you a discussion of *this* (*that*) crisis" (Vic-2). Thus, in his reading, FDR shifted these words to produce the focus of attention he desired or to indicate his approval or rejection of an idea.

Concern over clarity of ideas is again shown in the frequent oral repetition of a word or phrase before a modifying element. FDR read the word or phrase

in its place in the basic structure of the sentence, and repeated it spontaneously to form an anchor for the ensuing modification. Examples are numerous: "For what we send abroad, we shall be repaid, *repaid* within a reasonable time following the close of hostilities, *repaid* in similar materials" (FF-51); "And now we face another crisis—a *crisis* of a different kind but fundamentally even more grave" (Vic-2). And, further, when FDR introduced an impromptu modification of any significant length he often brought the framework of the sentence back into focus by repetition. Thus, FDR seemed intent upon holding the basic structure of the sentences clearly before his audience and ad-libbed to achieve this goal more fully.

Therefore, by compositional changes introduced at the moment of speaking, Roosevelt secured a text aimed at allowing the desired vocal emphasis and at providing clarity of ideas.

Roosevelt continued the tailoring of his text as he spoke; what had been so carefully prepared was altered in the speaking situation by changes of the same five types as had been made during the preparation of the Reading Copy. In the act of utterance Roosevelt was striving to clarify relationships still further, to render his coverage even more inclusive. He made few additions to add concreteness, but a very great many to add force to his communication and also a great many to heighten the responses of his listeners. He introduced compositional changes that served to allow him the vocal emphasis he desired and to stress the basic thought elements which he deemed necessary for making his meanings immediately and compellingly vivid to his listening audience.

ORAL COMMUNICATION AD-LIBS

But Roosevelt also made spontaneous changes directly associated with the act of oral communication. To remedy inadvertent, infrequent errors of reading and speaking, to prevent audience mis-hearing, to deal with audience responses—these purposes necessitated unforeseen alterations of a dozen kinds.

Corrections for bobbles

In some two and one-half hours of oral presentation (the five speeches combined) it might be expected that the speaker would experience some difficulty in mere accuracy of utterance, would get his "tongue twisted" a number of times. There are, however, but four such instances in the whole compass of these addresses. None occurred in the Pittsburgh address or the Four Freedoms speech, but either one or two appeared in each of the other speeches: "wide-sped *widespread* refusal" (Vic-56); "crystal clear *clear*" (MSG-26); "by the statesmen *statements* of men" (TU-18); and "such as that old warm *worm-eaten chestnut*" (TU-38). In all of these cases the correction was uttered with little if any break in rhythm and none in poise.¹¹

There are also a few examples of seemingly unintended repetition. He seemed to bounce on words at times: "I I asked the Congress" (MSG-33), and

¹¹ From his auditory study of the one address Voelker declared: "His feeling for rhythm in speech is quite exceptional. . . . His sense of rhythm is probably related to his sense of style. The rhythm is not lost when he makes mistakes. His corrections are exactly on the beat. That is a good criterion for judging how ingrained is his sense of rhythm. Furthermore, if the cadence is determined by the pace of a phrase before a pause, the first emphasized word after the pause is exactly on the beat. If the pause is lengthened by applause, this still holds true. Of course, the nature of the applause in the interval will vary the cadence. His perfection in this is such that it would seem that he had been keeping time during applause." *Loc. cit.*

"people themselves realize the *realize* the increasing urgency" (Vic-51). Elsewhere he seemed to have come sooner than he expected upon a word or phrase and the repetition was not to be prevented, as in "have been with us *with us*" (TU-17) and "to increase our national defense *defense*" (MSG-34). In one instance he repeated an unfinished word: "for a def- definite reduction of farm tenancy" (Vic-36). Though not bobbles, these are irregularities of vocal production and changes from the Reading Copy.

Correction of misreadings

One might also expect a considerable number of changes to correct misreadings; it is true, however, that FDR read with remarkable accuracy. There are a few cases of transposition of words:

from
who have already laid
from
I believe they are certainly among the ablest
men

and "candidates who burst out in election-year *election-year* affection for social legislation" (TU-24). In a few instances Roosevelt altered the text by anticipating later phraseology. For example, he replaced "we" with the words "you and I and everybody else" perhaps drawn from a sentence two paragraphs later (Pitts-5); later he substituted for "no one" the specification "nobody with any sense of responsibility," likely anticipating the idea of governmental responsibility explained four sentences later (Pitts-17).

Elsewhere he replaced immediately a misread word with the correct one, as in these instances: "the way to *ge-* win votes (MSG-37), "we oppose *forc-* enforced isolation" (FF-4), "factory owners and shop owners form *the background* the backbone" (Pitts-13), and "an or-

to who have laid already (TU-15); to I certainly believe that they are among the ablest men (MSG-12). ¹²

Infrequently he omitted a significant word and repeated the clause to pick it up, as in these instances:

from
And from where I stand it looks
from
These very men are now asking

to And where I stand and from where I stand it looks (Pitts-1); to These men these very men are now asking (TU-14).
--

On another occasion he enforced his correction with specific words:

from
When the first World War ended, we were

to Remember that when the World War, I mean the first World War, broke out, we were (MSG- 15).

There are several examples of misreading a word as a noun, then repeating it to make it the adjectival modifier it was meant to be: "And they too will resent this campaign *this campaign*-talk" (TU-17); "efficiency in defense *in defense* preparations of any kind" (FF-57);

derly peacetime *mobilization demobilization*" (TU-45). Two examples of misreading from the Teamsters' Union address are of more serious nature. In one

¹² These changes are considered as misreadings rather than purposeful alterations because repeated listening to the speech recordings reveals the passages as rather weakly read.

instance he hesitated, ad-libbed and repeated as though he had momentarily lost his place in the manuscript: "There was some—in the Congress and out—who raised their voices against our preparations for defense—before and after 1939, *objected to them, raised their voices against them*, as hysterical war mongering" (TU-13). In another instance he seemed to waver on a phrase as though he had lost the train of thought and then repeated to continue: "Our millions of soldiers and sailors and merchant seamen have been handicapped or prevented from voting by those politicians *prevented from voting by those politicians* and those candidates who . . ." (TU-26). These repetitions do not harm the meaning, and only their intonation reveals them as misreadings rather than as attempts to add force.

There are a few instances where the misreadings apparently escaped his notice: 1949 for 1941 (TU-13), 41 for 43 (Pitts-6), policy for problems (Vic-58), course for force (MSG-95), Congress for policy (MSG-31). Elsewhere he seems to have noticed his irregularity and showed by his pause that he thought it unnecessary to change, as when he read "perfectly" for "certainly" in this sentence, "The record is certainly clear" (MSG-43). On the other hand, he lost the meaning of a sentence when he misread "no" for "the" and seemed to weaken vocally as he continued without attempt to readjust: "There is no (the) task of finishing victoriously this most terrible of all wars" (TU-40). On another occasion he seems to have skipped a line; the Reading Copy had these words:

Without these spiritual forces we cannot make or maintain peace, and all three of them work with us toward that great end.

But FDR said this:

Without, without these three, all three of them, without their working with us toward

that great end, things would not be as clear or as easy (MSG-96).

It is interesting to note how often a misreading or a bobble is followed rather soon by another irregularity, either a mistake or an ad-lib. The Pittsburgh address has a misreading, then an ad-libbed caution against misunderstanding; the Victory Dinner speech has a bobble, then a misreading nine sentences after the bobble. In the Teamsters' Union address a bobble is followed by a completely ad-libbed sentence; there is also a bobble in the Fala paragraph (intensely dramatic and successful as that was), a misreading in the eighth sentence thereafter, then a spelling out in letters the word *botched* in the second sentence after that. Momentary faltering and a second error are reactions which one might expect in a speaker, but following an error with an ad-libbed element is behavior revealing remarkable confidence and skill. Even more noteworthy, however, is the extremely small number of errors of any type that occurred in these five important addresses.

Recovering from audience blurring of words

When audience response broke into a sentence, FDR often repeated enough of the sentence to provide a basis for understanding the remainder: "In fact, it looked so much like a shut-out for the team that you voted a change in management in order to give [applause] *in order to give* the country a chance to win the game" (Pitts-3). Further, he handily supplied words to make explicit a relationship that would have been clear had not the audience response blurred his words: "Not holding company executives or lawyers, [growing applause] *I'm talking about* men experienced in actual production" (MSG-11). Not many examples of such blurring

occurred, however, for Roosevelt was for the most part in charge of the applause of his audiences, calling it forth almost at will by the play of his voice, inviting vigorous response strongly and surely by subtle, sudden combinations of pitch, force and rate.

Taking pains to avoid mishearing

When Roosevelt spelled out the word *botched* ("They were botched—*b-o-t-c-h-e-d* they were botched by a Republican Administration" TU-43), he seems to have been ad-libbing more for emphatic effect and audience enjoyment than to avoid confusion in understanding. In another sentence, however, a direct reference to spelling does seem to have been aiming at clarity: "you and I—all of us together—were making forty-one billion—*spelt with a 'b' not an 'm'*—less dollars in 1932" (Pitts-6). There is little doubt even here, of course, but that Roosevelt was fully aware of the added satiric effect he was achieving.

Handling audience responses

FDR responded directly to the audience's eager reactions, making effective use of his need for restraining them: for example, he held back applause that was beginning, "*But wait, wait, it's been fifty-five years . . .*" (Pitts-35), and he halted the booing so that he could elicit an even greater response with his punch line, "*now wait, a perfectly beautiful alteration—Congressmen Martin, Barton and Fish*" (MSG-53).

He also ad-libbed to terminate audience responses. To conclude the vigorous applause which he had elicited by a pugnacious thrust he ad-libbed "and so" with a rising inflection and a commanding tone as he began the following sentence: "*And so on top of having to meet the ordinary expenses of the government . . .*" (Pitts-10). Shortly

thereafter he quieted the applause with "Yes, and I refused to leave human needs . . ." (Pitts-14). Sometimes he demanded their attention by repetition of an element present in the Reading Copy (as he bypassed boos for Hoover's administration by repeating *increased the debt*, Pitts-24) or of an ad-libbed one (as he acknowledged the cheers of his "old friends" but moved into the next statement, "*You know—you know, this is not the first time*" TU-1). Thus he often provided expendable words with which to quiet his audience so that, having utilized the effect of their interruption, he might then proceed.

These corrections for bobbles and misreadings, these additions to deal with audience blurring of his words, to hold back or terminate audience responses and to prevent listener misunderstanding—these ad-libbed elements are evidences of the interaction between vigorous audiences and a speaker of great sensitivity and skill.

It is clear then that two types of ad-libbing were occurring in these speeches: those alterations which continued the general processes of careful speech preparation, and those which arose from the speaker's immediate interaction with his listeners. It is interesting to note that Roosevelt was herein obeying certain well-known precepts of public speaking: (1) that the actual labor of creation should be going on vigorously during the act of speaking, the speaker seeking to express by word and voice the fullness of an idea, its relationships, its purpose; (2) that the speaker should, during the act of speaking, so impassion himself with his meaning that he seeks spiritedly for words that allow him suitable vocal and connotative play for his swelling thoughts; (3) that the speaker should, as he speaks, so involve himself with his listeners' responses that

he seeks actively for words that will keep before them vividly the ideas they must not miss, that will focus their attention and heighten their feelings.

Did these ad-libs alter greatly the message of the addresses? No, probably not;¹³ but they did alter greatly the *impact* of those messages, for they allowed Roosevelt to relate his ideas and himself to his listeners in more vibrant, compelling terms.

SPECULATION

Mindful of the careful labor of text-preparation in which FDR and his assistants regularly engaged, one feels impelled to ask: why did Roosevelt ad-lib so much? It is true that he as any other speaker had to recover from bobbles and misreadings; but when this has been considered it accounts for but a small percentage indeed of the total changes that FDR made in speaking from these carefully prepared texts. Most of the changes were voluntary and spontaneous.¹⁴ Why did he make them? The best answer seems to be: because he enjoyed it so much.

One might reasonably ask where FDR might have gained the ability to ad-lib so easily and confidently as to afford him such a measure of pure joy and amusement. Certainly his long hours daily on the telephone¹⁵ and in face-to-face conversation with men of widely differing outlook gave him vast experience. His frequent press conferences—

¹³ Rosenman pointed out that Roosevelt got himself into remarkably little trouble through them. *Op. cit.*, p. 486.

¹⁴ A somewhat different type of spontaneity sometimes occurred, however. Both Rosenman and Sherwood have pointed out that if FDR strongly wished to use a term or idea which his assistants vigorously opposed he might give in to its elimination from the text but hold it in mind and put it in at the time of delivery. Rosenman, 483-484; Sherwood, 218.

¹⁵ Gunther suggested that about one-fourth of his work day was spent in telephoning. John Gunther, *Roosevelt in Retrospect* (New York: Harper & Bros., 1950), p. 125.

nearly one thousand in all with 1 to 2 hundred questioners present¹⁶—gave him readiness with words and phrases and much opportunity to observe reactions to terms and tones at relatively close range. But these are matters of skill, and the true sources of FDR's enjoyment in ad-libbing seem to lie more deeply—in factors of personality, in characteristic attitudes. These three seem significant: his sensitivity to the forces about him; his love of experimentation, improvisation, change, mystery; and his pleasure in oral creativity.

Sensitivity

FDR understood and liked people, individually and in groups.¹⁷ He was able to assess their involvement and their degree of acceptance of his ideas and of himself; he was further able to gauge accurately what would heighten their feeling and give them further enjoyment. Constant practice over the years had sharpened this perceptiveness and his confidence in his power.

Love of experimentation

Roosevelt's "quicksilver mind,"¹⁸ his "nimbleness"¹⁹ tended to keep a play of alternatives before him. He loved to represent himself as a magician who could amaze and amuse his audience by "pulling another rabbit out of a hat."²⁰

¹⁶ "The President knew most of those who attended regularly on first-name or nickname terms; he was like a friendly, informal schoolmaster conducting a free-for-all seminar, and indeed the chief function of the conferences was educational. The President would explain an issue, take it apart, and present it over and over again from a new angle; he tackled the exposition of a difficult piece of legislation almost like an engineer demonstrating a new model of some machine." Gunther, pp. 135-136.

¹⁷ He was "easy in his way with crowds" (Burns, p. 281); he was "insatiably curious about people and their ideas, sensitively attuned to the play of forces around him" (*Ibid.*, p. 403).

¹⁸ *Ibid.*, p. 475.

¹⁹ Gunther, p. 114.

²⁰ Sherwood, p. 73. Gunther pointed out that he "loved to 'pull' surprises and put on a show." *Op. cit.*, p. 53.

This love of movement from the expected, this enjoyment of change, this delight in improvisation—these traits were deep in FDR. Burns has made the excellent suggestions that it was his "powerful sense of belonging" which gave him a sense of footing from which he was free to try the new, the never-before-dared frontier.²¹

Enjoyment of oral creativity

Winans pointed out, as we all remember, that the effective speaker must have *full realization of the content of his words as he utters them and a lively sense of communication*. Roosevelt not only fulfilled these obligations but deeply enjoyed performing them. He seemed to take great pleasure in making everything abundantly clear;²² he enjoyed the apt word, the memorable phrase; he knew instinctively the timing²³ and the tone that would give them most expressiveness. Indeed he thought of

²¹ *Op. cit.*, p. 78.

²² Burns explained that "His mind yearned for the detail, the particular, the specific." *Op. cit.*, p. 334.

²³ Gunther, pp. 236-237.

himself as an excellent actor, and truly he was.²⁴ His audience heard a man reacting vigorously to his ideas and inviting them powerfully to comprehend and be appropriately moved also, and no hearer could mistake his enjoyment of the whole process.

It is not hard to understand that such a man—sensitive to the feelings of the crowd and at home before them, experimental in his approach to ideas and thus only partially bound by the text before him, orally creative and delighting in the exercise of his talents—would have ad-libbed frequently and have enjoyed it hugely. Whether these alterations came as mere continuation of the usual efforts of his speech preparation or as a function of the interaction between himself and his audience, they were spontaneous adjustments to meet the realities of a situation by a speaker who not only perceived these realities with great clarity but who made consequent alteration with great ease and great enjoyment.

²⁴ Rosenman, p. 338; Gunther, p. 62; Burns, p. 474; Sherwood, p. 218.

AN EXPERIMENTAL INVESTIGATION OF RHETORICAL CLARITY*

ROGER E. NEBERGALL
University of Oklahoma

THIS paper is concerned with the application of experimental method to the study of one aspect of rhetorical style, variously called clarity, lucidity, or perspicuity. Writers on rhetorical theory from the time of Aristotle have usually agreed that the study of rhetorical style is an important part of the study of rhetoric, and that clarity is a fundamental quality or aspect of effective style. Little work has been done in the area of style by those interested in experimental research in communication. So, the general question, "Is it feasible to study rhetorical style by experimental method?", is more appropriate than the more specific question, "What are the elements of effective style?" An affirmative answer to the first question, of course, would suggest that later research could profitably be concerned with the second.

Specific questions posed by this study in an effort to find out whether style, or more particularly clarity, can be investigated by the procedures used here are:

1. Do people differ in their ability to construct "clear" messages?
2. Is a message "clearer" in written form than it is in oral form?
3. Is the ability to construct a "clear" message related to the general verbal ability of an individual?
4. Will messages judged "clear" by the measures here proposed also be judged "clear" by rhetorical critics?
5. Will messages judged "clear" by the measures here proposed be judged "comprehensible" by readability techniques?

*Based on a Ph.D. dissertation, University of Illinois, 1956, directed by Marie K. Hochmuth.

In order to answer these questions, a definition of clarity is needed which will make measurement of this quality possible. The following definition of clarity, which is related to the definitions proposed by many rhetorical theorists, is proposed:

A message is clear to the degree that a speaker's intended meaning for a message agrees with the audience's obtained meaning from the message.

A completely clear message, according to this definition, would be one from which the audience obtained exactly the same meaning as the speaker intended for the message, and the clarity of the message would be less than complete to the extent that the meaning which an audience obtained from a message differed from the meaning that a speaker intended for it. The experimental application of this definition presumes, of course, an objective measure of meaning. The measure of meaning employed in this study is Osgood's semantic differential.¹

Since the semantic differential is regarded by those who developed it as a logic, rather than as a test in any final form,² it was necessary to develop a form of the semantic differential to employ in the experiment proposed in this

¹ The most complete discussion of the semantic differential is in Charles E. Osgood, George J. Suci, and Percy H. Tannenbaum, *The Measurement of Meaning* (Urbana: The University of Illinois Press, 1957).

² Charles E. Osgood, "Report on the Development and Application of the Semantic Differential" (Unpublished Monograph, University of Illinois, Institute of Communications Research), p. 14.

TABLE I*
THE TWENTY-FOUR SCALES EMPLOYED IN THE EXPERIMENT AND THE FACTOR LOADING
ON EACH OF THE FACTORS.

	Factor I (Evaluation)	Factor II (Potency)	Factor III (Activity)
1. good-bad	.88	.05	-.09
2. large-small	.06	.62	.34
3. beautiful-ugly	.86	.09	.01
4. hard-soft	-.48	.55	.16
5. sweet-sour	.83	-.14	-.09
6. strong-weak	.19	.62	.20
7. calm-agitated	.61	.00	-.36
8. red-green	-.33	-.08	.35
9. loud-soft	-.39	.44	.23
10. deep-shallow	.27	.46	.14
11. sharp-dull	.23	.07	.52
12. ferocious-peaceful	-.69	.17	.41
13. heavy-light	-.36	.62	-.11
14. hot-cold	-.04	-.06	.46
15. nice-awful	.87	-.08	.19
16. treble-bass	.33	-.47	.06
17. angular-rounded	-.17	.08	.43
18. fragrant-foul	.84	-.04	-.11
19. honest-dishonest	.85	.07	-.02
20. active-passive	.14	.04	.59
21. fair-unfair	.83	.08	-.07
22. rugged-delicate	-.42	.60	.26
23. fast-slow	.01	.00	.70
24. cleandirty	.82	-.05	.03

*From Charles E. Osgood, "Report on the Development and Application of the Semantic Differential" (Unpublished Monograph, University of Illinois, Institute of Communications Research), p. 9.

study. This was done by selecting the eight sets of polar adjectives showing the highest factor loadings on each of the three factors which have been so far identified. The twenty-four sets of polar adjectives employed, together with their factor loadings are shown in Table I.

A form of the semantic differential, when filled out by a subject who checks one of the seven possible positions on the seven-point scale separating each of the sets of polar adjectives employed, generates a profile of the subject's meaning for the particular concept tested. In the present experiment, the concern is with the relationship of two such profiles, the profile generated by the speaker, and the profile generated by a member of an audience. These profiles can be compared by the D method developed by

Osgood and Suci.³ The D formula is:

$$D = \sqrt{\sum d^2}$$

In this formula, d represents the distance between the two profiles on each of the scales. On a form of the semantic differential employing 24 scales, D has a minimum value of 0 and a maximum value of $\sqrt{864}$ or 29.39, since the largest possible d on each scale is 6 making $d^2 = 36$, and

$$24 \cdot 36 = 864.$$

PROCEDURE

Reproductions of five abstract paintings on slides suitable for projection were secured, and five students were selected to serve as communicators. Two of the communicators selected were

³ Charles E. Osgood and George J. Suci, "A Measure of Relation Determined by Both Mean Difference and Profile Information," *Psychological Bulletin*, XXXIX (May, 1952), 251-62.

graduate students in speech at the University of Oklahoma; the other three were undergraduate students at the same institution. In the experiment, they were asked to describe the meaning they perceived in the abstract pictures. Abstract paintings were selected for this purpose since it was felt that the use of such stimuli would insure that the messages would provide as much as possible of the available information about the stimulus for the members of the audiences employed.

Each communicator was shown each of the pictures (projected on a screen by a slide projector) and asked to explain his meaning for it, "so that anyone hearing your explanation will understand the picture just as you understand it." Two or three days later, he was shown the pictures again and was asked to record his meaning for them on the semantic differential. With the limitation that each picture was presented to one of the communicators first, to another second, etc., the pictures were presented in random order. The second presentation of the pictures was ordered the same as the first.

The messages obtained in this fashion were literally transcribed. Appropriate punctuation was inserted and random vocalizations omitted.

Ten audiences of 20-25 persons each, consisting of sections of the beginning speech course at the University of Oklahoma, were arbitrarily divided into two groups of five audiences each. One group heard the recorded versions of the messages, and the other read the messages when they were presented in written form.

Two Latin squares were randomly drawn, governing the order in which the messages would be presented to each audience. The procedure followed with

each audience for the recorded versions of the messages was as follows:

A tape was prepared with the five messages that the particular audience was to hear. The instructor of that class introduced the experimenter, and the experimenter instructed the group in the procedure of the experiment and the use of the scales of the semantic differential. The first message was played on a tape recorder and the audience filled out the semantic differential for that message. Then the next message was played and the procedure repeated. The total time required to administer the test for the five messages was about 25 minutes.

A similar procedure was used for administering the test with the written form of the message except that the printed copy of the message was distributed and was available while the subjects were differentiating their meaning for it. Thus, for the oral form of the message, the subjects had to recall the message as they were filling out the scales of the semantic differential; for the written form, they could refer to the message while filling out the scales. It was felt that this provided a realistic comparison of the oral and written forms, since one of the fundamental differences between oral and written communication is that a reader has an opportunity to re-read the communication while the listener has an opportunity to hear the communication only once.

Various other measures were then gathered. Scores for all audience members on the Ohio State Psychological Examination were obtained. Also, twelve rhetoricians were asked to judge the messages as to their "clarity." This was done by asking them to record their judgment as to the clarity of the message by checking the appropriate space on a

seven point scale ranging from "extremely clear," to "extremely unclear."

Finally the messages were evaluated as to their "comprehensibility" by subjecting them to "cloze" analysis. The "cloze procedure" is a measure of the comprehensibility of messages, developed by Wilson Taylor which correlates highly with the standard readability measures while appearing to avoid some of the most obvious defects in them.⁴ For the "cloze" analysis, student groups comparable to the original audiences were employed.

ANALYSIS OF DATA

The following hypotheses were tested in the analysis of the data obtained in the experiment:

1. That there is a difference among the communicators employed in their ability to communicate their intended meaning.
2. That there is a difference in the general level of transmission of meaning between the oral and written forms of the messages.
3. That the differences among the messages in their effectiveness in conveying a speaker's

intended meaning to an audience are related to the differences in clarity of the same messages as judged by rhetorical critics.

4. That the ability of a member of an audience to obtain a speaker's intended meaning is related to his verbal ability as measured by the Ohio State Psychological Examination.
5. That the differences among the messages in their effectiveness in conveying a speaker's intended meaning to an audience is related to the differences in comprehensibility of the same messages as determined by "cloze procedure."

Differences in Ability to Communicate Intended Meaning

This hypothesis was tested by analysis of variance of the data obtained for both the oral and written forms of the messages. The cell entries in Table II are the means of the D scores obtained between the speaker's meaning profile for the message, and the profile for the message of each member of the audience, for the audiences which heard the oral form of the messages. Table III shows the results of the calculation

TABLE II
MEAN AUDIENCE D.

Speaker	Oral Messages					Sum
	I	II	III	IV	V	
1	11.31 (Mc)	11.12 (Md)	12.86 (Mb)	11.09 (Ma)	13.35 (Me)	59.73
2	11.12 (Md)	11.66 (Mc)	15.85 (Ma)	13.93 (Me)	10.96 (Mb)	63.52
3	13.59 (Mb)	11.95 (Me)	14.10 (Mc)	17.72 (Md)	14.56 (Ma)	71.92
4	12.78 (Ma)	10.23 (Mb)	13.11 (Me)	11.24 (Mc)	14.64 (Md)	62.00
5	10.04 (Me)	8.31 (Ma)	10.27 (Md)	10.00 (Mb)	9.96 (Mc)	48.58
Sum	58.84	53.27	66.19	63.98	63.47	/305.75

Note: Mx identifies the message generated by the picture (x) when presented to the communicator in the same row.

⁴ Wilson Taylor, "'Cloze Procedure': A New Tool for Measuring Readability," *Journalism Quarterly*, XXX (Fall, 1953), 415-433; also "Recent Developments in the Use of 'Cloze Procedure,'" *Journalism Quarterly*, XXXIII (Winter, 1956), 42-48.

of analysis of variance for the data in Table II. Table IV shows the data for the five audiences which received the written form of the messages. Table V

TABLE III
ANALYSIS OF VARIANCE FOR ORAL MESSAGES.

Sources of Variation	df	Mean Square	F
Columns (Audiences)	4	5.31	2.46
Rows (Speakers)	4	14.12	6.54*
Treatments (Pictures)	4	1.90	—
Residual	12	2.16	

*Significant at the 1% level of confidence.

their intended meaning to the audiences employed.

The Oral and Written Forms

The significant difference between audiences evidenced in the analysis of variance applied to the written form of the message raises some question about the nature of the parent population (all

TABLE IV
MEAN AUDIENCE D.

Speaker	Written Messages					Sum
	I	II	III	IV	V	
1	13.57 (Mc)	10.29 (Mc)	10.08 (Ma)	13.74 (Mb)	10.87 (Md)	58.55
2	9.74 (Md)	11.87 (Me)	11.88 (Mc)	15.82 (Ma)	11.64 (Mb)	60.95
3	13.22 (Mc)	13.60 (Ma)	13.81 (Mb)	17.54 (Md)	11.32 (Me)	69.49
4	12.50 (Ma)	10.35 (Mb)	13.29 (Md)	12.76 (Me)	12.11 (Mc)	61.01
5	8.76 (Mb)	9.52 (Md)	9.27 (Me)	8.92 (Mc)	9.45 (Ma)	45.92
Sum	57.79	55.63	58.33	68.78	55.39	/295.92

shows the results of the calculation of analysis of variance for the data in Table IV. It will be noticed that in both forms of the experiment the null hypothesis that there is no difference

possible sections of the beginning speech course at the University of Oklahoma). For this reason, it was decided to compare the results of the oral and written forms of the messages by a non-parametric statistic which makes no assumptions about the nature of the parent population.

Applying a sign test⁵ to the speaker totals in Tables II and IV we find that for all five speakers the total D score is greater in the oral form of the test than in the written. Using a two-tailed test of significance, this difference is significant at about the 2 per cent level.

Applying the same sign test to the individual message scores in the two tables, however, we find that seventeen

TABLE V
ANALYSIS OF VARIANCE FOR WRITTEN MESSAGES

Sources of Variation	df	Mean Square	F
Columns (Audiences)	4	6.09	3.42*
Rows (Speakers)	4	14.45	8.12**
Treatments (Pictures)	4	3.21	—
Residual	12	1.78	

*Significant at the 5 percent level of confidence.

**Significant at the 1 percent level of confidence.

between communicators can be rejected ($p < .01$), and therefore it can be concluded that there was a significant difference between the communicators employed in their ability to communicate

⁵ Helen Walker and Joseph Lev, *Statistical Inference* (New York: Henry Holt & Company, 1953), p. 430.

of the D scores are higher for the oral form of the test, and eight are higher for the written form of the test. Using a two-tailed test of significance, this difference is not significant.

These results are probably not sufficiently conclusive to enable us to reject the null hypothesis, and to conclude that the speakers' meanings were better transmitted by the written form of the messages than by the oral form. At the same time they suggest that a more precise analysis might show significant differences between the oral and written forms of the messages.

Rhetorician's Judgments of Message Clarity

In testing the third hypothesis, the sum of D scores for each message in both the oral and written forms of the test was obtained. This sum of D scores was regarded as a measure of the clarity of the message. Using this score, the messages were rank-ordered as to their clarity. The judgments of the rhetorical critics on the seven point rating scale were summed for messages, with a check to the far left of the scale being recorded as a 1 and a check to the far right of the scale being recorded as a 7. These total scores were then rank-ordered. The rank-orders assigned the messages by the two methods were then compared by rank-order correlation. The rank-order correlation obtained was -.33 which is not significantly greater than could be expected by chance.

From these results we may conclude that we are unable to reject the null hypothesis, and unable to show that the differences among the messages in their effectiveness in conveying a speaker's intended meaning to an audience are related to the clarity of the same messages as judged by rhetorical critics.

O.S.P.E. Scores

The fourth hypothesis was tested by analysis of variance. The question was whether the D scores of individuals in different O.S.P.E. deciles could be regarded as random samples from the same population, or whether the mean D for individuals in one O.S.P.E. decile might differ from the mean D for individuals in another O.S.P.E. decile. Two O.S.P.E. scores (total and reading) and two sets of individual D scores (written and oral) were available. Four analyses of variance were computed (all possible combinations). The results of these analyses of variance are reported in Tables VI, VII, VIII, and IX.

TABLE VI
ANALYSIS OF VARIANCE OF INDIVIDUAL D SCORES
AND O.S.P.E. TOTAL, ORAL FORM.

Source of Variation	df	Mean Square	F
Between groups	9	47.85	—
Within groups	79	56.96	
Total	88		

TABLE VII
ANALYSIS OF VARIANCE OF INDIVIDUAL D SCORES
AND O.S.P.E. READING, ORAL FORM.

Source of Variation	df	Mean Square	F
Between groups	9	133.00	2.84*
Within groups	79	46.74	
Total	88		

*Significant at the 1 percent level of confidence.

TABLE VIII
ANALYSIS OF VARIANCE OF INDIVIDUAL D SCORES
AND O.S.P.E. TOTAL, WRITTEN FORM.

Source of Variation	df	Mean Square	F
Between groups	9	42.26	—
Within groups	49	52.93	
Total	58		

TABLE IX
ANALYSIS OF VARIANCE OF INDIVIDUAL D SCORES
AND O.S.P.E. READING, WRITTEN FORM.

Source of Variation	df	Mean Square	F
Between groups	9	20.33	—
Within groups	49	56.95	
Total	58		

In these four analyses of variance, the results differ significantly from chance in only one. In the analysis involving the D scores obtained from audience members hearing the oral form of the message, and the O.S.P.E. Reading scores, a difference significant at the 1 per cent level is noted. Therefore, we may conclude that for the oral form of the messages tested here, there is a significant relationship between the ability of an audience member to obtain a speaker's intended meaning and his verbal ability as judged by his reading score on the Ohio State Psychological Examination.

Cloze Scores

The fifth hypothesis was tested by rank order correlation. The cloze scores obtained for each message were used as a basis for ranking the messages as to their comprehensibility. The rank-order correlation between these ranks and the ranks obtained for the clarity of the messages in testing Hypothesis III was computed. The rank order correlation was +.26 which is not significant.

Thus we may conclude that we cannot reject the null hypothesis and are unable to establish that there is a relationship between the comprehensibility of a message as determined by the cloze formula, and the clarity of the same message as judged by the methods employed here.

DISCUSSION

Interpretation of Communicator Differences

The nature of the Latin square design is such that while the presentations of column variables, row variables and treatment variables can be randomized in presentation, the order of presentation of the row variable cannot be randomized, and therefore the results ob-

tained for the row variable are confounded with the order of presentation. For this experiment, this means that while each audience heard one message from each speaker which had been selected at random (with the limitation that each message be heard by one audience), and one message about each picture (with the limitation that a message about a certain picture would be presented first to one audience, second to another, etc.), each audience heard a message from the same speaker first, from the same speaker second, etc.

While this is a limitation of this particular design, there is no reason to believe that this confounding affected the results obtained. Studies that have been done on the effects of speaker order have reached somewhat ambiguous conclusions, but their conclusions have not indicated that the results obtained by this experiment could have been predicted on the basis of speaker order. These studies have indicated that speaker positions other than first and last tend to be favored,⁶ so the results in this experiment, with the lowest D scores obtained for the first and last speakers would indicate that this type of order preference was not operating in this case.

The Evaluative Dimension

As will be seen by reference to Table I, the form of the semantic differential employed in this experiment is rather heavily weighted on the evaluative dimension. This is in spite of the fact that the twenty four scales selected consisted of the eight scales with the highest factor loadings on each of the three dimensions. This is due, of course, to

⁶ F. H. Knower, "A Study of Rank Order Methods of Evaluating Performance in Speech Contests," *Journal of Applied Psychology*, XXIV (October, 1940), 633-644; also, Sam L. Becker, "The Ordinal Position Effect," *Quarterly Journal of Speech*, XXXIX (April, 1953), 217-219.

TABLE X
D SCORES—SPEAKER PROFILE AND AUDIENCE MEDIAN PROFILE.

Speaker	Oral Form Audience					Sum
	I	II	III	IV	V	
1	7.21 (Mc)	7.21 (Md)	6.56 (Mb)	5.66 (Ma)	8.43 (Me)	35.07
2	6.24 (Md)	5.74 (Mc)	8.72 (Ma)	7.87 (Me)	4.36 (Mb)	32.93
3	8.18 (Mb)	6.40 (Me)	8.66 (Mc)	10.63 (Md)	5.57 (Ma)	39.44
4	7.00 (Ma)	5.29 (Mb)	5.83 (Me)	6.71 (Mc)	7.35 (Md)	32.18
5	5.38 (Me)	1.73 (Ma)	4.47 (Md)	5.20 (Mb)	4.00 (Mc)	20.78
Sum	34.01	26.37	34.24	36.07	29.71	/160.40

the fact that the factor loadings on the evaluative factor are considerably higher than the factor loadings on the other two dimensions, and also to the fact that evaluative elements are found in many of the activity and potency scales.

TABLE XI
ANALYSIS OF VARIANCE OF
DATA-MEDIAN PROFILES.

Source of Variation	df	Mean Square	F
Columns (audiences)	4	3.10	2.46
Rows (speakers)	4	9.77	7.75*
Treatments (pictures)	4	1.32	—
Residual	12	1.26	—

*Significant at the 1% level of confidence.

Reference to the text of the messages employed in this experiment indicates that in some cases small profile differences existed between speaker and audience for messages which appeared to consist largely of evaluative statements. It could be argued that the differences observed between the speakers was not due to differences in ability to communicate their intended connotative meaning, but to differences in the evaluative content of the messages.

To evaluate this possibility, a nine-scale form of the semantic differential was employed. These nine scales included the three activity and potency scales with the highest loadings, and

TABLE XII
D SCORES—SPEAKER PROFILE AND AUDIENCE MEDIAN PROFILE.

Speaker	Written Form Audience					Sum
	I	II	III	IV	V	
1	7.81 (Me)	6.00 (Mc)	5.66 (Ma)	6.40 (Mb)	6.00 (Md)	31.87
2	6.71 (Md)	7.48 (Me)	6.40 (Mc)	9.16 (Ma)	5.00 (Mb)	34.75
3	8.37 (Mc)	6.16 (Ma)	9.27 (Mb)	10.00 (Md)	5.66 (Me)	39.46
4	6.78 (Ma)	4.47 (Mb)	6.78 (Md)	5.47 (Me)	6.32 (Mc)	29.82
5	4.90 (Mb)	4.12 (Md)	3.87 (Me)	4.79 (Mc)	1.73 (Ma)	19.41
Sum	34.57	28.23	31.98	35.82	24.71	/155.31

produced a test which was less heavily weighted on the evaluative dimension. The same analysis of variance that was originally performed was repeated with

TABLE XIII
ANALYSIS OF VARIANCE OF
DATA-MEDIAN PROFILES.

Written Form			
Source of Variation	df	Mean Square	F
Columns (audiences)	4	4.19	2.87
Rows (speakers)	4	11.45	7.84*
Treatments (pictures)	4	.53	—
Residual	12	1.46	

*Significant at the 1% level.

this new data. The results of this analysis are reproduced in Tables X, XI, XII, XIII.

As can be seen from this analysis, some minor changes were produced, but

the significance of the communicator differences remains unchanged, and the relative position of the speakers with the highest and lowest D scores remains unchanged. Apparently the differences originally noticed were not due to an excessive loading of the original instrument on the evaluative dimension.

To further explore the possibility that the differences between the speakers were due largely to judgments made on the evaluative dimension, another analysis was made employing the five of the nine scales that have almost no loading on the evaluative dimension. These five scales were the strong-weak, large-small, hot-cold, active-passive, and fast-slow scales. The results of these calculations are reported in Tables XIV, XV, XVI, and XVII.

TABLE XIV
D SCORES—SPEAKER PROFILE AND AUDIENCE MEDIAN PROFILE
ACTIVITY AND POTENCY SCALES.

Speaker	Oral Form					Sum
	Audience					
1	5.83 (Mc)	6.85 (Md)	3.16 (Mb)	5.29 (Ma)	6.32 (Me)	27.45
2	5.48 (Md)	3.60 (Mc)	6.40 (Ma)	4.90 (Me)	3.16 (Mb)	23.54
3	7.28 (Mb)	4.47 (Me)	6.08 (Mc)	8.66 (Md)	5.00 (Ma)	31.49
4	4.00 (Ma)	3.16 (Mb)	5.10 (Me)	4.69 (Mc)	3.16 (Md)	20.11
5	2.64 (Me)	1.41 (Ma)	4.24 (Md)	4.90 (Mb)	3.87 (Mc)	17.06
Sum	25.23	19.49	24.98	28.44	21.51	119.65

TABLE XV
ANALYSIS OF VARIANCE OF DATA-MEDIAN PROFILES
ACTIVITY AND POTENCY SCALES.

Oral Form			
Source of Variation	df	Mean Square	F
Columns (audiences)	4	2.29	—
Rows (speakers)	4	6.69	3.62*
Treatments (pictures)	4	1.55	—
Residual	12	1.85	

*Significant at the 5% level of confidence.

It will be noted that the results of the original analysis change very little even with this violent departure from the original conditions of investigation. The differences between the communicators are still significant, although only at the 5 per cent level of confidence. Apparently, the activity and potency scales enable us to distinguish between the messages just about as well by them-

TABLE XVI
D SCORES—SPEAKER PROFILE AND AUDIENCE MEDIAN PROFILE
ACTIVITY AND POTENCY SCALES.

Speaker	Written Form					Sum
	Audience					
I	II	III	IV	V		
1	6.56 (Me)	5.20 (Mc)	5.09 (Ma)	3.32 (Mb)	6.00 (Md)	26.17
2	5.92 (Md)	5.10 (Me)	3.74 (Mc)	6.56 (Ma)	3.16 (Mb)	24.48
3	6.48 (Mc)	5.00 (Ma)	8.48 (Mb)	7.07 (Md)	4.36 (Me)	31.39
4	4.47 (Ma)	2.64 (Mb)	3.16 (Md)	5.10 (Me)	4.12 (Mc)	19.49
5	4.59 (Mb)	3.74 (Md)	2.50 (Me)	3.87 (Mc)	1.41 (Ma)	16.21
Sum	28.12	21.68	22.97	25.92	19.05	/117.74

TABLE XVII
ANALYSIS OF VARIANCE OF DATA-MEDIAN PROFILES
ACTIVITY AND POTENCY SCALES.

Written Form			
Source of Variation	df	Mean Square	F
Columns (audiences)	4	2.44	—
Rows (speakers)	4	7.13	3.58*
Treatments (pictures)	4	.31	—
Residual	12	1.99	—

*Significant at the 5% level of confidence.

selves as when the evaluative scales are added. It should also be noted that this analysis produces the highest total D for speaker 3 and the lowest D total for speaker 5 as did the two preceding analyses.

We may conclude, therefore, that the distinctions between the speakers discovered in the original investigation are not due solely to an overloading of the scales on the evaluative dimension, since repeated analyses which remove more and more of the evaluative factor merely operate to diminish the differences and not to cause them to disappear.

The Ohio State Psychological Examination

In considering the relationships between the scores obtained by the mem-

bers of the audiences on the O.S.P.E. and their total D scores, two primary limitations must be considered. First, there was no reason related to the design or hypotheses of the experiment for the selection of the O.S.P.E. instead of any other test of verbal ability. The reasons for the choice were simply availability and convenience. The Ohio State Psychological Examination is administered as part of the freshman testing program at the University of Oklahoma, and the scores obtained on the examination were available for the members of the audiences participating in the experiment. For this reason, it might not be as useful as some testing instrument selected with the design and hypotheses of the experiment in mind.

The second limitation regarding the O.S.P.E. scores is that scores on the examination were not available for all of the students serving as members of the audiences for the experiment, since transfer students to the University of Oklahoma are not required to take the freshman test battery. Thus scores were not available for those members of the audiences who were transfer students. Out of a total of 206 students who were audience members in this experiment,

O.S.P.E. scores were unavailable for 60. However, there seemed to be no reason to expect these omissions of O.S.P.E. scores to be other than random in relation to the experiment, and therefore relationships were simply calculated utilizing the scores available.

Of the four analyses of variance which were calculated, a significant relationship appeared in only one. Since there were no *a priori* reasons to expect significant relationships in certain combinations and not in others, it would not be sound to select the largest F and argue that it represents the most highly significant relationship, especially since the four analyses of variance indicated only one significant F. However, this F is significant at better than the 1 percent level, which argues fairly strongly that this particular analysis of the data indicates a real relationship and is not merely capitalizing on chance. However, because of this limitation, the significant F in the analysis of variance between the O.S.P.E. reading score and the total D scores on the oral form of the test should be regarded as indicative but not conclusive.

The Judgments of the Rhetorical Critics

The lack of relationship between the D scores obtained for the various messages and the judgment of these messages as to their clarity by rhetorical critics needs to be further explored. Clarity is defined in this study as the relationship between a speaker's intended meaning and an audience's obtained meaning, and the D scores obtained are intended as a measure of this relationship. The only information the rhetorical critics had as to the intended meaning of the speaker was such information as could be obtained from inspection of the messages. From an inspection of the messages, the critics had

to determine the speaker's intended meaning, and decide how effectively this particular message would communicate that meaning to an audience. If the measure proposed in this study is valid, we were unable to demonstrate that the rhetorical critics were able to do this. Actually, this failure is not too surprising. It seems doubtful that there is any necessary relationship between the form and content of a message and the speaker's intended meaning. Since the definition of clarity proposed here agrees with definitions proposed by rhetorical theorists, this lack of relationship would appear to be a criticism, not of the critics, but of the definitions of clarity which appear in traditional rhetorical theory.

The "Cloze Procedure"

The "cloze procedure" is intended as a measure of the "comprehensibility" of a message. It would be expected that the comprehensibility of a message would be a factor in its clarity, or in other words, the "understandability" of a message generated by a speaker would be a factor in how well that message would communicate a speaker's intended meaning to an audience. On the other hand, it would not be expected that the comprehensibility of a message would determine or fix its clarity. From these considerations, one would predict that the clarity of a message and its comprehensibility would be positively correlated, but that the correlation would be fairly low. The correlation obtained ($R = +.26$) is not significant at the 5 per cent level, but a correlation this size would be obtained between unrelated variables by chance only about once in ten times. It would thus appear that while this experiment is unable to establish a significant relationship between "clarity" as here defined and

"comprehensibility" as defined by the "cloze procedure," further exploration of this relationship is worthwhile.

SUMMARY

This study has investigated a particular aspect of rhetorical style, the aspect of clarity. Clarity is defined as the relationship between a speaker's intended meaning for a message and the audience's obtained meaning from the message. The semantic differential is employed as an instrument to measure these meanings, and the D score is employed as a measure of the extent of their difference.

The results of the experiment indicated that the communicators did vary significantly in their ability to convey

their intended meaning to an audience; that the written form of the messages may have been more effective, and that the study was unable to relate the communication of a speaker's intended meaning to clarity as defined by rhetorical critics.

It was also found that the ability of a member of an audience to obtain a speaker's meaning in oral presentation was related to his reading score on the O.S.P.E.

Finally, it was found that there was no apparent relationship between clarity as defined here, and comprehensibility as defined by the "cloze procedure." While the obtained correlation was positive, it was not significantly larger than could be expected by chance.

A POSSIBLE BASIS FOR THE ASSOCIATION OF VOICE CHARACTERISTICS AND PERSONALITY TRAITS

EDITH B. MALLORY and VIRGINIA R. MILLER
Wellesley College

I. BACKGROUND

THERE is a widespread tendency for listeners to attribute specific personality characteristics to an individual on the basis of his speech. Audiences often derive their impressions of a lecturer from the sounds of his voice as well as from the words that he utters. It is customary for personnel managers to be influenced in their decisions by the voices of those they interview. In the informal conversations of daily life the majority allow their judgments of persons to be affected by cues of this sort. In all these cases there is evidently an implicit assumption that the voice is, to some degree, a reflection of the personality. Psychologists recognize this possibility, and include various aspects of speech among the "expressive movements"¹ which they regard as vehicles of personality projection.

In experimental investigations the relationships between voice and personality have been most often approached by attempting to determine whether or not certain personality characteristics can be discerned from the *total* voice pattern. Kelly² found that personality ratings by acquaintances agreed more closely with self-ratings than did those made by radio listeners who had never seen the speakers. Allport and Cantril³

found the relationships between the scores on a number of personality tests and estimates of personality characteristics as judged from the voice. Judgments based on the voice were compared with objective test scores for dominance by Eisenberg and Zalowitz,⁴ and by Fay and Middleton for introversion,⁵ sociability,⁶ and Spranger Personality Types.⁷

A number of investigators have been concerned with the degree to which effectiveness in speech correlates with specific personality characteristics. Gilkinson and Knower^{8,9} studied the relationship between the ratings of effectiveness in speech and scores on the Bell Adjustment Inventory and other tests. Chenoweth¹⁰ reported Bernreuter scores and

¹ P. Eisenberg and E. Zalowitz, "Judging Expressive Movement: III. Judgments of Dominance-Feeling from Phonographic Records of Voice," *Journal of Applied Psychology*, 22 (1938), pp. 620-631.

² P. J. Fay and W. C. Middleton, "Judgment of Introversion From the Transcribed Voice," *Quarterly Journal of Speech*, 28 (1942), pp. 226-228.

³ P. J. Fay and W. C. Middleton, "The Ability to Judge Sociability from Voice As Transmitted Over a Public Address System," *Journal of Social Psychology*, 13 (1941), pp. 303-309.

⁴ P. J. Fay and W. C. Middleton, "Judgment of Spranger Personality Types From the Voice As Transmitted Over a Public Address System," *Character and Personality*, 8 (1939), pp. 144-155.

⁵ Howard Gilkinson and Franklin H. Knower, "Individual Differences Among Students of Speech As Revealed by Psychological Tests," *Quarterly Journal of Speech*, 26 (1930), pp. 243-255.

⁶ Howard Gilkinson and Franklin H. Knower, "A Study of Standardized Personality Tests and Skills in Speech," *Journal of Educational Psychology*, 32 (1941), pp. 161-175.

⁷ Eugene C. Chenoweth, "The Adjustment of College Freshmen to the Speaking Situation," *Quarterly Journal of Speech*, 26 (1940), pp. 585-588.

¹ G. W. Allport and P. E. Vernon, *Studies in Expressive Movement* (New York, 1933).

² E. L. Kelly, "Personality As Revealed by Voice and Conversation Without Face to Face Acquaintance," *Psychological Bulletin*, 35 (1938), p. 710 (Abstract).

³ G. W. Allport and H. Cantril, "Judging Personality from Voice," *Journal of Social Psychology*, 5 (1934), pp. 37-55.

case history data of effective and non-effective speakers. Dow¹¹ also studied personality test results as related to effectiveness in public speaking. Further studies in this area have been made by Duncan,¹² Eckert and Keyes,¹³ Moore,^{14, 15} and Murray.^{16, 17}

Since even the total voice pattern is only an approximate indication of personality, one cannot suppose that close relationships would be found between personality traits and partial aspects of speech. Sapir¹⁸ discussed the possibility of such relationships but offered no experimental or quantitative evidence to confirm their reality. Moses,^{19, 20} also, analyzed various aspects of voice and suggested their connection with psychological conditions. Duncan²¹ reported significant correlations of quality, pitch, and force with scores on the Bell Inventory and the Minnesota Personality Scale by Darley and McNamara. She

found only non-significant relationships between Bernreuter scores and students' ratings of these voice factors, and so excluded the Bernreuter results from her final investigation. Moore,²² however, used this test as well as self-ratings in an investigation of students whose voices were breathy, harsh or metallic or had a nasal whine. Sanford^{23, 24} found indications of personality characteristics in particular aspects of speech, but his analysis was primarily concerned with features such as literary style and vocabulary, i.e., at levels more complex than voice qualities.

The writers of this paper are interested not so much in identifying speech characteristics as symptoms of personality conditions, but rather in the possibility of formulating a basic principle which might account for the relationships which do exist. In general, investigators whose aim has been to find diagnostic clues to personal adjustment have discovered that it is easier to judge personality from total speech patterns than from any of the simpler, analyzed factors such as pitch or rate. Allport, for instance, says, "To attempt to correlate pitch with one personal quality, speech with another, and intensity with a third would be to make the whole problem absurdly atomistic, and, as is the case with all studies which seek correlations between mere meaningless fragments of well-structured personalities, the study would be foredoomed to failure."²⁵ This will be so if one is concerned primarily with individual prognoses, but in an inquiry designed to discover a tenable basis for the relationship between

¹¹ Clyde W. Dow, "The Personality Traits of Effective Public Speakers," *Quarterly Journal of Speech*, 27 (1941), pp. 525-532.

¹² Melba Hurd Duncan, "An Experimental Study of Some of the Relationships Between Voice and Personality Among Students of Speech," *Speech Monographs*, 12 (1945), pp. 47-60.

¹³ R. G. Eckert and N. Keyes, "Public Speaking As a Cue to Personality Adjustment," *Journal of Applied Psychology*, 24 (1940), pp. 144-153.

¹⁴ Wilbur E. Moore, "Personality Traits and Voice Quality Deficiencies," *Journal of Speech Disorders*, 4 (1939), pp. 33-36.

¹⁵ Wilbur E. Moore, "Factors Related to Achievement and Improvement in Public Speaking," *Quarterly Journal of Speech*, 29 (1943), pp. 213-217.

¹⁶ Elwood Murray, *The Speech Personality* (Philadelphia, 1937), chapt. 6.

¹⁷ Elwood Murray, "A Study of Factors Contributing to Mal-Development of the Speech Personality," *Speech Monographs*, 3 (1936), pp. 99-108.

¹⁸ E. Sapir, "Speech As a Personality Trait," *American Journal of Sociology*, 32 (1927), pp. 892-905.

¹⁹ Paul J. Moses, "Social Adjustment and Voice," *Quarterly Journal of Speech*, 27 (1941), pp. 532-537.

²⁰ Paul J. Moses, *The Voice of Neurosis* (New York, 1954).

²¹ Duncan, *op. cit.*

²² Wilbur E. Moore, "Personality Traits and Voice Quality Deficiencies," pp. 33-36.

²³ F. H. Sanford, "Speech and Personality: A Comparative Case Study," *Character and Personality*, 10 (1942), pp. 169-198.

²⁴ F. H. Sanford, "Speech and Personality," *Psychological Bulletin*, 39 (1942), pp. 811-845.

²⁵ Allport and Cantril, *op. cit.*, p. 39.

speech and personality and the procedure which Allport rejects may actually be indicated. A fundamental pattern of relationships which occurs with significantly greater than chance frequency may be of theoretical value, even though the correlations which are found prove to be far too low to warrant clinical determinations.

II. PROBLEM AND HYPOTHESIS

Insofar as any expressive movements are indications of phases of personality, we should ask how such forms of reactions may have become established. It is not genuinely explanatory to say that a particular subjective trait "makes us" act in a particular way, especially in the case of physical habits over which we at times exercise little deliberate control. Some voice habits may be largely dependent on physical structure whereas others may have been developed chiefly through imitation or another type of learning. It is the writers' hypothesis that some vocal habits which are more directly associated with personality traits were developed from muscular responses to the very same stimuli which originally induced the subjective states typical of those traits. Thus, both the motor habits and the personality traits related to them would have been established by the same sequence of earlier events.

According to this description of the possible connection between subjective traits and related expressive movements, the components of an individual's voice pattern would be the audible aspects of certain physical responses which have become habitual in social situations involving communication with other persons. Some concomitance should therefore be discoverable between those personality characteristics and those voice characteristics which

can be conceived as parallel relations to particular kinds of social stimuli.

In order to test this hypothesis, two personality variables were selected—dominance as measured by the B4D Scale of the Bernreuter Inventory, and Introversion as measured by the B3I Scale of the same test. Social dominance may be thought of as varying quantitatively, in different individuals, along a linear continuum which extends from high ascendancy to submission. The first extreme represents a highly positive attitude toward effectiveness in contacts with persons. As such it leads to habitually outgoing tendencies to action, especially in those facets of behavior which concern communication with the world of other individuals. Submission, on the contrary, is associated with habits of acceptance and restraint.

A series of events (or experiences) which would induce the subjective attitude of submission might include rebuffs or frustrations in situations involving other persons. A parallel physical response to such stimuli could be the inhibition of the actions incipient at the time of the event; thus, specific restraint of the actions normally employed in communicating with other persons.

We might therefore expect that submission would be accompanied by a tendency toward the tightening of the vocal cords and result in giving the voice a *higher pitch*. It might also cause a constriction of structures of the air passages, producing a tendency toward *lack of resonance*. The curtailment of vigorous action in the speech process might result also in a *reduction of loudness*. Introversion, the habit of directing attention toward the self rather than toward the outside world, is also often related to difficulties in making out-

side personal contacts. Insofar as this is the case, introversion and submission would tend to be concomitant, and this is indeed a usual, but not invariable, finding. We might therefore anticipate that introversion would also tend to be associated with those voice characteristics which could be expected to accompany submission.

Since tension can lead to a rapid, jerky form of activity when compensatory effort is made, there might also be some likelihood of a *more rapid rate* in the speech of those who are found at the submissive end of the personality scale. In the case of introversion, however, which may involve the habit of ignoring the outside world rather than submitting to it, there would necessarily be a reason for the compensatory effort resulting in rapid rate. In accordance with this hypothesis we would therefore predict that dominance scores would be positively correlated with loudness, resonance, lower pitch and slower rate. We would predict also that introversion scores would correlate negatively with loudness, resonance, and lowness of pitch, but would not show any connection with rate.

III. THE EXPERIMENTAL INVESTIGATION

Procedure. The subjects of this investigation were all the members of an entering class at Wellesley College, a group relatively unselected in respect to speech and personality traits studied. Individual tests were given by the Speech Department, in which judgments were made of the voice of each student as she read aloud and spoke in a standard testing situation. Since the primary object of the examination was to identify those who might need to attend special speech classes, particular attention was paid to voice characteristics which might be construed as "de-

fects"; and therefore there was no step-wise rating of a quality such as pitch or rate, but note was taken of those cases in which the former was especially high or the latter was exceptionally rapid. Notations were made also of other outstanding characteristics, both desirable and undesirable, in descriptive words such as "hard," "resonant" and the like. In addition, the judges indicated those instances in which the manner of speaking was clearly "confident" or "self-conscious." It should be made quite plain that these latter terms were not attempts at general personality judgments, but were intended to define the impression made by the voice itself.

During the first week of their stay at college, the Bernreuter Personality Inventory was administered as a group test to these students. It was recognized that no questionnaire is an adequate measure of personality trends, but since chance errors in such scores would reduce, rather than enhance, the extent of any relationships that might be found, it was felt that these scores offered conservative, as well as objective, criteria for comparison. To facilitate statistical treatment, the Dominance score (B_4D) and the Introversion score (B_3I) were converted into T scores based on the Wellesley norms for the test. For these, the mean was 50 and the standard deviation 10.

Foreign students whose English speech was not properly comparable to that of the other students were excluded from the study, and no use was made of the partial records of the few individuals who missed one or the other test. Complete records were available for 372 subjects.

Results. Biserial correlations were found between each of the two personality measures, Dominance and Introversion, and the incidence of the

BISERIAL CORRELATIONS OF VOICE CHARACTERISTICS AND BERNREUTER SCORES N = 372

	Dominance				Introversion			
	r_{bis}	σr_{bis}	CR	P	r_{bis}	σr_{bis}	CR	P
High Pitch	-.15	.08	1.87	.03	.13	.07	1.85	.03
Inadequate Loudness	-.32	.07	4.57	.00	.14	.08	1.75	.04
Resonance	.14	.08	1.75	.04	-.24	.07	3.43	<.01
Rapid Rate	-.17	.08	2.12	.02	.00	—	—	—
Confident Manner	.28	.10	2.80	<.01	-.22	.07	3.14	<.01

voice characteristics recorded. The results found in the accompanying table show the degrees of positive or negative relationships as indicated by the biserial correlation coefficients (r_{bis}).²⁶ With each coefficient is given its standard error (σr_{bis}), and, in the CR column, the value found by dividing r_{bis} by σr_{bis} . The column headed P reports the level of confidence for each coefficient, i.e., the number of chances per hundred that the calculated relationship would occur merely by chance.

High pitch was found to correlate with a coefficient of $-.15$ with Dominance, and $.13$ with Introversion. These relationships are extremely slight, but obtain in the expected directions, at the 3 per cent level of confidence. Inadequate loudness shows, with Dominance, a correlation of $-.32$. This also is low, but since it is four and one-half times its standard error, it must be interpreted as very probably due to causative factors other than chance. Inadequate loudness and Introversion are less definitely related, with a coefficient of $.14$ at the 4 per cent level. Resonance

gives a correlation of $.14$ at the 4 per cent level with Dominance, and $-.24$, at the 1 per cent level, with Introversion.

It was suggested that rapid rate might be a concomitant of Submission and so show a negative relationship with Dominance scores, and this is, perhaps, shown by the coefficient of $-.17$ at the 2 per cent level of confidence. No correlation whatever appeared between rapid rate and introversion.

While the specific object of this investigation was the testing of our hypothesis regarding specific voice variables, it is interesting and not irrelevant to note also our findings regarding the personality measures in relation to the judgments of confidence versus self-conscious manner of speech. Confident manner correlated $.28$ with Dominance and $-.22$ with Introversion; and both of these coefficients show a relationship at the 1 per cent level of confidence. While these figures were not high, they exceed the majority of those found in connection with the isolated voice qualities. This corroborates the finding of other experimenters that an observer's judgment tends to be based not on isolated aspects, but on the total pattern presented. It is of course possible that the bearing of the student, i.e., the visible manner of standing and the facial expression may also have contributed to the verdicts. It appears that characteristics of personality reveal themselves more clearly in the complex of

²⁶ For High Pitch, Inadequate Loudness and Resonance, the biserial correlations were found by using Formula 99, Garrett. Henry E. Garrett, *Statistics in Psychology and Education* (New York, 1937), p. 369. In the case of Rate and Manner, judgments made possible the use of groups categorized at the opposite extremes of the scale, i.e., rapid versus slow, and confident versus self-conscious. In these cases Formula 206, Peters and Van Voorhis, was used. Charles C. Peters and Walter R. Van Voorhis, *Statistical Procedures and Their Mathematical Bases* (New York, 1940), p. 385.

a variety of expressive movements than they do in the isolated aspects of voice production, although the latter may be components of the total effect.

Discussion. All of the correlations found in this study are low. They offer no justification for predicting an individual's personality characteristics from isolated voice characteristics, or for predicting the latter from the former. For the purposes of this study, however, the essential question is not "how close are the relationships," but "do any of the expected correlations exist to an extent that shows they are due to causative factors other than chance?" While only two of the specific voice characteristics showed correlations on which specific voice characteristics showed correlations on which reliance can be placed at better than the 1 per cent level of confidence, all of them are found to be at the 4 per cent level, or better. The findings offer appreciable evidence that certain personality and voice characteristics are definitely associated, in ways which support our hypothesis that the two types of responses may have been conditioned concomitantly by events in the individual's earlier life history.

It may reasonably be asked why, if simultaneous parallel conditioning occurs at all, the association found between the personality traits and the voice characteristics is so slight. Probably the answer lies in the fact that the human being is an animal highly susceptible to social or cultural influences. Speech is learned largely by imitation, and a large majority of one's habits of speech may be picked up in the same

way. Moreover, there is considerable conscious control of the voice, so that many modifications may be made deliberately in the direction of socially approved standards. Also, since it is widely understood that certain idiosyncrasies will be interpreted unfavorably by one's associates, a compensatory effort to correct or mask revealing inadequacies will be made,—either intentionally or unconsciously, even by the introverted and submissive individual. In view of the extensive overlay of modifications so acquired, it is of considerable significance that the pattern of relationships which supports the hypothesis of this study has been found to persist.

IV. CONCLUSIONS

1. There is slight but significant positive association between the personality trait of dominance and the voice characteristics of loudness, resonance, and lower pitch. Submission by contrast has a slight but negative association with the voice characteristics of loudness, resonance, and lower pitch.
2. There is a slight positive association between submissiveness and rapid rate of speech.
3. Introversion is negatively related to loudness, low pitch and resonance, but is unrelated to rate.
4. The pattern of relationships found in this study supports the hypothesis that certain personality traits and certain voice characteristics may have developed concomitantly as persisting parallel reactions to particular kinds of situations involving social communication.

PERSONALITY AND DISCUSSION BEHAVIOR: A STUDY OF POSSIBLE RELATIONSHIPS

THOMAS M. SCHEIDEL, and LAURA CROWELL

*University of Washington
and*

JOHN R. SHEPHERD
University of Oregon

IN recent years a number of factor analytic studies have been carried out in the hope of isolating some essential dimensions of individual behavior in small groups. The concurrence among the various researchers regarding the identity of the important factors is remarkable. As summarized by Carter,¹ the three basic dimensions are:

FACTOR I: *Individual Prominence and Achievement*

These are behaviors of the individual related to his efforts to stand out from others and individually achieve various personal goals.

FACTOR II: *Aiding Attainment by the Group*

These are behaviors of the individual related to his efforts to assist the group in achieving goals toward which the group is oriented.

FACTOR III: *Sociability*

These are behaviors of the individual related to efforts to establish and maintain cordial and socially satisfying relations with other group members.

As an exploratory study from which more intensive investigations are projected, the research reported here was intended to pursue the relationships between personality value and need structures and effectiveness in group discussion. It was also hoped that this research would provide some workable hypotheses about the relations of personality and the three dimensions of behavior described by Carter.

At first glance, this may appear to be

a "trait" approach to the study of leadership, an approach which has been roundly and soundly criticized. Most would agree that the simple listing and tabulating of the traits correlated with leadership is a "sterile approach." As Gibbs writes:

... viewed in relation to the individual, leadership is not an attribute of the personality but a quality of his role within a particular and specified social system. Viewed in relation to the group, leadership is a quality of its structure.

.....
Leadership is both a function of the social situation and a function of personality, but it is a function of these two in interaction.²

It is undoubtedly true that in our search for some consistent personality configurations a field interpretation is needed, involving a *person* in a *situation*. The importance of the *situation* in producing the leader must never be slighted. As in Barrie's play *The Admirable Crichton*, the model butler's leadership qualities come to light only after his whole party is marooned on a deserted island. But neither can those interested in discussion discount the importance of the *person*. In a narrowly defined area, such as small group problem-solving discussion, leadership certainly is not entirely situationally defined. If there exist some paradigms of personality which can be consistently re-

¹ See Carter's summary of research leading to these factors in: Launor F. Carter, "Recording and Evaluating the Performance of Individuals as Members of Small Groups," *Personnel Psychology*, VII (1954), 477-484.

² Cecil A. Gibb, "The Principles and Traits of Leadership," *Journal of Abnormal and Social Psychology*, XLVII (1947), 267-268.

lated to leadership behaviors, they must be fully investigated. A tabulation of leader traits is not so much needed as studies of the dynamics of personality in small groups. The necessity for such investigation is mentioned by Gibb as he discusses *self-assurance* as a trait of leadership:

A person who believes in himself gives the impression that he has the skill, power, or ability which will enable him to solve the problem in hand.

This is one of those personality traits which makes it abundantly clear that individual personality cannot be left out of the leadership picture. Leadership cannot be exclusively a function of the situation as it is seen by an independent observer, for individual differences clearly affect the social perceptions of some individuals by others, and consequently play an important part in giving structure to the situation for those who are a part of it.³

With these considerations in mind this exploratory study of the personality needs and values related to effectiveness in discussion groups has been completed. It has been the purpose of this research to produce some hypotheses regarding the dynamics of personality as it relates to the observable dimensions of small group behavior.

I. PROCEDURE

Sixty-six undergraduate students from four sections of the basic group discussion course at the University of Washington served as the subjects for this study. Each of the subjects completed three measurement devices during the quarter. These included two personality inventories, the Allport-Vernon-Lindzey *Study of Values*⁴ and the Edwards *Personal Preference Schedule*.⁵ In addition,

³ Cecil A. Gibb, "Leadership," *Handbook of Social Psychology*, Volume II, edited by Gardner Lindzey, (Cambridge, Mass., 1954), 886.

⁴ Gordon V. Allport, Philip E. Vernon, and Gardner Lindzey, *Study of Values, A Scale for Measuring the Dominant Interests in Personality*, (Boston, 1951).

⁵ Allen L. Edwards, *Edwards Personal Preference Schedule*, (New York, 1954).

each student completed the *Discussion Rating Scale* prepared by the authors.

1. The Allport-Vernon-Lindzey *Study of Values*. This test purports to measure personality in terms of the values that the individual considers important in life. The term "values," as used here, can be equated with "interests" or "motives," and these values are based upon those identified by Spranger.⁶ The six variables and the *loci* of their dominant interest are:

1. Theoretical—observation, reasoning and systematization of knowledge;
2. Economic—usefulness, practicality;
3. Aesthetic—gracefulness, symmetry, fitness;
4. Social—sympathy, kindness, unselfishness;
5. Political—personal power, influence, renown;
6. Religious—personally satisfying relationships with the universe.

In this scale the respondent is asked to indicate his preference for alternative answers to forty-five questions; on the basis of his responses the relative preference for each of these values may be determined.

2. The *Edwards Personal Preference Schedule*. This schedule measures fifteen personality variables, derived from those proposed as "manifest needs" by Murray.⁷ Edwards tests these needs with a schedule containing 225 pairs of statements. The subjects are to choose that statement in each pair which they believe to be the more characteristic of themselves. These scores provide a measure of the following personality needs:

1. Achievement—need to be successful, to accomplish significant results;
2. Deference—need to accept the leadership of others;
3. Order—need to obtain and maintain organization and system;
4. Exhibition—need to receive the notice and comment of others;

⁶ Eduard Spranger, *Types of Men*, trans. by Paul J. W. Pigors, (Halle, Germany, 1928), 109-246.

⁷ H. A. Murray, et. al., *Explorations in Personality*, (New York, 1938), 76-85.

5. Autonomy—need to be free to act and to criticize;
6. Affiliation—need to build and maintain attachment to others;
7. Intraception—need to analyze motives and predict behaviors;
8. Succorance—need to receive sympathy, encouragement and favors from others;
9. Dominance—need to influence and supervise the actions of others;
10. Abasement—need to feel guilt, blame and inferiority to others;
11. Nurturance—need to treat others with kindness, sympathy and generosity;
12. Change—need to experience novelty;
13. Endurance—need to work hard, persistently and to the completion of the task;
14. Heterosexuality—need to have pleasant relations with members of the opposite sex;
15. Aggression—need to attack others and to avoid personal blame.

3. *The Discussion Rating Scale.* This instrument, prepared by the authors, is a fifteen step rating scale which provides a basis for locating each subject along a continuum of discussion effectiveness. The number of scoring categories employed in this scale was based on the discussion of rating scales by Guilford.⁸ Each student was provided with a copy of the rating form on which had been listed the names of the class members. He was instructed to rate each student in his class on the basis of his total impression of that student's discussion effectiveness, having heard the student take part a number of times during the term either as member or appointed leader of a discussion group. He was reminded that the person's effectiveness involved his behavior both in the logical development of ideas and in the maintenance of group atmosphere conducive to a high calibre of participation by all.

A brief statement of the nature of the basic discussion course in which these ratings were made is pertinent

here. Each student participated in a small group discussing a topic before the class a half dozen times during the term, acting as designated leader at least once. Concepts of effective membership and leadership were evolved in the class through frequent and detailed evaluations of observed behavior. Responsibilities of members for voluntarily filling both logical and psychological needs arising in the group were clarified, and responsibility of the leader for stimulating and focusing the activity of the group toward its goal was commonly understood. That these responsibilities are not deemed the singular province of either the designated leader or the members makes possible a composite judgment.

One further step was taken in the preparation for this research. Prior to the final analysis of the data, the authors attempted to predict the tendency of the effective discussants on some of the key variables. Assuming personality needs and values to be related to the three dimensions of discussion behavior, the authors predicted certain relationships. These predictions and the results are included as Table III and discussed later in this paper.

II. STATISTICAL ANALYSIS

1. *Selection of Groups.* A scale value, or position on the continuum of discussion effectiveness, was determined for each subject by applying the method of equal-appearing intervals to the rating scale data. Following a rank ordering of the subjects according to these values, those discussants who fell in the upper or lower 27% of their class were selected. By combining discussants thus selected from the four sections, it was possible to form groups of 18 relatively effective discussants and 18 relatively ineffective discussants.

⁸ J. P. Guilford, *Psychometric Methods*, 2nd Ed., (New York, 1954), 263-301.

2. *Statistical Procedures.* The raw scores for each subject on both personality scales were converted to T scores, with a mean of 50 and a standard deviation of 10, by referring to the college population norms included in the test manuals. The appropriate sex norms were consulted in order that the group differences tested would not be influenced by the sex variable. To test the hypothesis that there are no significant differences between effective and ineffective discussants, the standard t test for independent groups was applied. In order to test the hypothesis that this group does not differ from the normative college populations, it was necessary to evaluate the statistical significance of the difference found between the means of the standardization groups, which may be regarded as population means, and the means for the effective discussants. The t test was used for this purpose.

III. RESULTS

The results of the analyses described above are summarized in Tables I, II, and III. Table I contains the data regarding the differences between the relatively effective and relatively ineffective discussants. Two significant differences were found. The effective discussants were significantly lower on Affiliation and higher on Dominance than the ineffective discussants. It is possible to generalize, on the basis of Murray's observations and on Edwards' list of items, that the individual scoring low on Affiliation would not have a need to feel a sense of belongingness; he would not be especially interested in forming new friendships or maintaining old ones and would not place a high premium on social contact in general. A person scoring high on Dominance could be characterized as possessing a strong personal need

to be a leader, to be recognized by a group as the leader, to be individually prominent in a group.

Table II summarizes the data regarding the differences between effective discussants and the normative college populations. This comparison reveals three statistically significant differences. The effective discussants are significantly higher on Dominance and lower on Succorance and Abasement than the normative populations. The Succorance variable can be considered synonymous with the term "support" as used in psychological terminology. The individual who scores low on this variable would feel little need for increased encouragement and affection from others. He would not have a primary need to have others show interest, sympathy, and understanding toward his personal problems. He would not need the support of others. A low score on Abasement suggests that one has little feeling of timidity, inferiority, or self-depreciation. Of especial interest are two variables on which the effective discussants did not deviate significantly. On the Allport Social value and the Edwards Affiliation need, contrary to the expectations of statistically significant high rankings, this group ranked below the mean of the normative populations.

In summary, these data would suggest that the effective discussants can be characterized by a need to be recognized as leader, to be independent, to be self-confident, and by a lack of concern for social contact and affiliation. These subjects evince a high value in and personality need for self-sufficiency and personal influence.

Table III summarizes the results of the predictions made by the authors regarding the relationships between the personality variables and the dimensions of discussion behavior. Such prediction

TABLE I
DIFFERENCES BETWEEN EFFECTIVE AND INEFFECTIVE DISCUSSANTS.

Variable	Effective Discussants Mean N=18	Ineffective Discussants Mean N=18	Mean Difference	t
<i>Study of Values</i>				
theoretical	53.17	50.56	2.61	.92
economic	46.67	53.00	6.33	1.97
aesthetic	47.33	45.50	1.83	.65
social	48.22	45.72	2.50	.81
political	51.56	54.17	2.61	.88
religious	52.33	52.06	.27	.08
<i>Personal Preference Schedule</i>				
achievement	52.06	51.94	.12	.04
deference	50.11	54.61	4.50	1.14
order	50.56	52.67	2.11	.67
exhibition	48.39	49.44	1.05	.25
autonomy	49.33	44.83	4.50	1.36
affiliation	47.72	54.56	6.84	2.33*
introception	52.11	49.28	2.83	.92
succorance	44.44	49.22	4.78	1.54
dominance	59.56	49.78	9.78	3.32**
abasement	44.11	49.83	5.72	1.54
nurturance	49.78	52.17	2.39	.78
change	48.33	50.83	2.50	.78
endurance	52.33	48.78	3.55	1.00
heterosexuality	48.28	46.83	1.45	.45
aggression	52.61	46.22	6.39	1.97

*Significant at 5% level of confidence.

**Significant at 1% level of confidence.

TABLE II
DIFFERENCES BETWEEN EFFECTIVE DISCUSSANTS
AND NORMATIVE COLLEGE GROUPS.

Variable	Effective Discussants Mean N = 18	t
<i>Study of Values</i>		
theoretical	53.17	1.58
economic	46.67	1.22
aesthetic	47.33	1.96
social	48.22	.81
political	51.56	.75
religious	52.33	1.00
<i>Personal Preference Schedule</i>		
achievement	52.06	1.32
deference	50.11	.04
order	50.56	.23
exhibition	48.39	.58
autonomy	49.33	.29
affiliation	47.72	1.23
introception	52.11	.96
succorance	44.44	2.30*
dominance	59.56	5.06**
abasement	44.11	2.80*
nurturance	49.78	.09
change	48.33	1.06
endurance	52.33	1.19
heterosexuality	48.28	.77
aggression	52.61	1.27

*Significant at 5% level of confidence.

**Significant at 1% level of confidence.

was difficult because the categories of personality are not unequivocal and because some of these traits overlapped the behavioral factors. Reference to the trait definitions given earlier should help clarify for the reader the reasoning behind the predictions, although a few of the predictions may require further explanation. High aggression was not predicted for the effective discussants because the aggression variable in the Edwards test would not seem to contribute to the type of individual prominence that discussion theorists would consider conducive to leadership. For the same reason it was predicted that this group would score low on the Exhibition variable. A high Economic value was predicted for it was considered that a stress on the "practical" and the "useful" would contribute to group goal facilitation behavior.

The interesting results of this anal-

TABLE III
A PRIORI PREDICTIONS OF AUTHORS REGARDING CERTAIN PERSONALITY TRAITS
RELATED TO THE THREE DIMENSIONS OF DISCUSSION BEHAVIOR.

INDIVIDUAL PROMINENCE	GROUP GOAL FACILITATION	GROUP SOCIABILITY
high theoretical*	high theoretical*	high social
high political*	high economic	high affiliation
high achievement*	high achievement*	high nurturance
low exhibition*	high order	
high autonomy	high endurance*	
high dominance*		
low abasement*		
high endurance*		

*Those predictions which seem largely to be borne out in terms of the group deviation from the normative mean by at least one unit.

ysis were: 1. The relationships hypothesized between certain personality variables and "Individual Prominence" were found in the group of effective discussants. 2. The relationships hypothesized between certain personality variables and "Group Goal Facilitation" and "Group Sociability," with the exception of those traits overlapping "Individual Prominence," were not found in the group of effective discussants.

IV. INTERPRETATION AND DISCUSSION

It was the purpose of this exploratory study to provide hypotheses for more intensive investigations. The most provocative hypothesis suggested by the data of this project can be stated as follows: *Certain personality need and value structures can be directly related to behaviors of individual prominence in small group discussions; such structures cannot be consistently related to behaviors of group goal facilitation and group sociability for these latter factors involve more fully the interactional elements of the situation.*

Some discussion of this suggested hypothesis may be helpful. It is undoubtedly true that there is no unitary trait of leadership. Yet, in narrowly defined situations such as small group discussion, certain persons do seem consistently to provide most of the leadership behavior. When we compare these ef-

fective discussants with the average group member or ineffective group member, we do find some significant personality differences. The finding with the greatest statistical significance in this study is the high personality need for Dominance evinced by the effective discussants. The literature on this point has not been consistent. Krech and Crutchfield have held that "other things being equal, those persons who have inconsistent needs for dominance, power, and prestige may be expected to have higher potentiality for leadership."⁹ On the other hand, Cattell and Stice¹⁰ have recently found no significant differentiation between leaders and non-leaders in terms of dominance. Gibb, in the face of the previous contradictory studies, writes, "Perhaps the best that may be said at the moment is that there is no clear evidence of the dependence of leadership upon dominance-need."¹¹ It would seem that the results of this study support the view of Krech and Crutchfield and stand in opposition to the position of Gibb.

The results of this study also lead to the view that the effective discussant typically has a personality need for self-

⁹ D. Krech and R. S. Crutchfield, *Theory and Problems of Social Psychology*, (New York, 1948), 437.

¹⁰ See R. B. Cattell and G. F. Stice, *The Psychodynamics of Small Groups*, (Urbana, 1953).

¹¹ Gibb, "Leadership," *op. cit.* p. 888.

confidence and independence. The first factor is not surprising, for most relevant studies in the literature find leaders to be self-confident. Regarding the second factor, a number of studies report the leader traits of initiative, persistence, ambition. Gibb writes,

One is constrained to point out here that these findings do not, in fact, support the frequently heard contention that leaders are people who force themselves by repeated efforts into positions of leadership. Rather, an interpretation more consistent with the data would be that this trait of "will" and "persistence" is highly prized by groups which, in facing problems, have previously met failure and rebuff.¹²

Not only in groups facing opposition, but in many other groups as well, a leader exhibiting the characteristic of independence may find approval.

A surprising lack of need for sociability was demonstrated by the experimental group. The literature on this point, as on dominance, is contradictory, some studies showing the social trait and others not. The present study, like the research of Richardson and Hanawalt,¹³ did not demonstrate any significant need by the effective discussants for sociability. They were below the mean on all variables related to this factor. An explanatory hypothesis would be that effective discussants have the "capacity" or "ability" for social behavior but not a personality "value" in or "need" for it.

In this study of the dynamics of personality in a small discussion group an attempt has been made to consider the possible interaction of *personality* and *situation*. The behavioral factor of Individual Prominence seems to have some basis in certain of the personality variables studied. More research designed to determine relevant traits and to tap

the relationships of these traits with this first dimension of behavior is needed. Of course, the ultimate test of the products of such studies will be their value in predicting discussion behavior. The behavioral factors of Group Goal Facilitation and Group Sociability could not in this study be shown to have such a basis in personality structure. It is hypothesized, therefore, that these dimensions of discussion behavior are products of certain aspects of the situation rather than directly of personality. Social perception and group interaction studies should clarify relationships among these factors and add to our understanding of this aspect of communication.

V. SUMMARY

This study proposed to examine certain personality needs and values in relation to performance in small group discussions. Sixty-six students in basic discussion classes rated each other as to their overall effectiveness as group discussion participants. Analysis of these ratings in relation to data provided by the *Edwards Personal Preference Schedule* and the *Allport Study of Values* suggested notable relationships between such personal characteristics as self-confidence, independence and dominance and the "Individual Prominence" dimension of discussion behavior. The analysis, however, revealed no such relationships of personal characteristics with the other two behavioral dimensions, "Group Goal Facilitation" and "Group Sociability." These results suggest this interesting hypothesis: certain personality need and value structures can be directly related to behaviors of individual prominence in small group discussions but such structures cannot be consistently related to behaviors of group goal facilitation and group sociability.

¹² *Ibid.*, p. 887.

¹³ H. M. Richardson and N. G. Hanawalt, "Leadership as Related to the Bernreuter Personality Measures," *Journal of Social Psychology*, XVII (1943), 237-267.

EXPERIMENTAL STUDIES OF MOTIVATED GROUP DISCUSSION*

WILLIAM E. UTTERBACK
WALLACE C. FOTHERINGHAM
Ohio State University

EXPERIMENTAL research in conference method encounters two principal difficulties: motivating subjects to feel a concern for the success of the discussion and measuring the quality of the group product. Both relate to the choice of subject matter for discussion in the experimental situation. If the group's task, for example, is to solve an intellectual puzzle or estimate the number of beans in a bottle, the quality of the group product is measurable but the subject may have little interest in the outcome. Substituting a more realistic task, such as formulating policy on a controversial public question, makes measurement difficult without significantly improving motivation. Both realistically motivated discussion and measurable outcomes are essential to the experimenter's purpose. The studies here reported attempt to avoid these difficulties in an investigation of several hypotheses related to conference method.

BASIC METHOD

The group task. Discussion in the experimental situation was of a series of hypothetical problems dealing with discussion method (Experiments 1 and 2) and with ethics and human relations in industry (Experiments 3 and 4). Each problem, stated in a short paragraph, was followed by a list of five solutions, the group's task being to select the best solution. Of the five, one, the "pre-

ferred" solution, was known by the experimenter to be the best in the judgment both of a panel of subject-matter experts and of a majority of the approximately 100 students who had previously discussed the problem in an undergraduate course in discussion method. The problems had been pre-validated by use as practice material in the course. It was thus known that at least 30 per cent but not more than 60 per cent of the students in the course choose the preferred solutions before discussion and that discussion moves opinion toward those solutions. It was known also that the different series of problems used in each experiment were of approximately equal difficulty.

Subjects and procedures. Subjects were undergraduates enrolled in sections of a course in discussion method in which practice discussion of problems similar to those used in the experiments was a familiar part of course routine. In both practice and experimental sessions the subject was given a specified length of time to study each problem and to record on a signed ballot his pre-discussion choice of best solution. After discussion of the problem for a specified length of time each subject recorded his post-discussion choice. The two sets of ballots from the experimental sessions provided the data for the studies.

Previous to both practice and experimental sessions subjects received the following advice:

Each of you will receive three grades on your performance in these discussions. The first will

*This study was supported in part by funds granted to the Ohio State University for aid in fundamental research by the Research Foundation of Ohio State University.

be based on the proportion of times you choose the preferred solution before discussion, for you are learning how to study a problem before undertaking discussion of it; the second will be based on the proportion of times you choose the preferred solution after discussion, for you are learning how to profit from discussion; the third will be based on the percentage of the members of your groups who choose the preferred solution after discussion, for you are learning how to help the group reach as high a degree of consensus as possible on a sound solution.

Subjects were thus motivated to take the experimental situation seriously, to exercise sound judgment both before and after discussion, and to give the group maximum assistance in choosing the best solution.

Measures of the outcomes of discussion. Three outcomes of discussion were employed as criterion variables. The first was the amount of shift of opinion occurring in discussion, measured by the number of shifts recorded on pre- and post-discussion ballots from one choice of solution to another, regardless of the correctness of the choice. The second was progress toward consensus, measured without reference to correctness of choice by the Matthews Index of Agreement, which yields scores ranging from zero to one.¹ The third was the amount of improvement in quality of judgment effected by discussion, the criterion of quality being the keyed best solution, as indicated above. As it was necessary to vary the method of computing this measure somewhat from one experiment to another, the specific method employed in each case is explained later.

Experimental variables. Three variables were under investigation: the length of discussion, varied from two to

13 minutes for each problem; the size of the group, varied from three to 12; and the style of moderation.²

Three styles of moderation were employed. Under the first no moderator was designated in advance, though nothing was done to prevent one member of the group from emerging spontaneously as moderator, and this undoubtedly happened in some discussions. Under the second style, "partial" moderation, a moderator was designated in advance but restricted his activities to regulating participation so that all members of the group had an equal opportunity to speak. Under the third style, "full" moderation, the moderator was instructed to facilitate group thinking, not only by regulating participation, but also by making suggestions regarding analysis of the problem, by asking questions to stimulate thinking, by making transitional and summary remarks, and by any other means that occurred to him as likely to be useful. Except as noted later, moderators did not vote before or after the discussion and were instructed not to express any opinion on the merits of any solution or of any argument presented in discussion or to do anything else to influence the group toward selection of a particular solution. In no case did the moderator have advance information regarding which was the preferred solution of any problem.

With one exception to be noted later, all moderators were drawn from an advanced course in conference leadership and had received about five weeks of training and experience in moderating

¹ Jack Matthews and A. W. Bendig, "The Index of Agreement: A Possible Criterion for Measuring the Outcome of Group Discussion," *Speech Monographs*, XXII (1955), 39-42.

² Here and throughout this paper the terms *moderator* and *moderation* are preferred to *leader* and *leadership* to avoid the implication which the latter terms carry in much of the literature on conference method that the person designated to facilitate group thinking may, or should, direct group thinking toward a particular conclusion.

discussion. In addition moderators drawn from this course had been put through a short period of intensive training in moderating discussion of such problems as were to be used in the experiment and had been tested to insure that they could follow instructions regarding the style of moderation to be employed.

Control of variables. When not under investigation, length of discussion, size of group, and style of moderation were either held constant or controlled through the experimental design. In addition, except as indicated later, the following variables were either held constant or controlled: length of subjects' previous training and experience in discussion, subjects' previous acquaintance with the designated moderator, subjects' previous knowledge of the problems discussed, length of moderators' previous training and experience, the various sections of the course, each taught by a different instructor, from which subjects were drawn, the different successive days on which the experiment was conducted (not controlled in the last two experiments, as no practice effect had been noted in the first two), the different series of problems used in each experiment, the competence of the individual moderators employed.

Except as indicated later for Experiment 4, the subjects drawn from each section of the course in each experiment were selected randomly and for the first day of the experiment were assigned randomly to experimental groups. On each succeeding day the same subjects were reassigned to groups in such a way as to minimize the number of times a subject encountered the same classmate in a group and to equalize the number of such encounters for the various experimental conditions.

EXPERIMENT 1. STYLE OF MODERATION

The purpose of the experiment was to determine whether two styles of moderation (no designated moderator and full moderation) differ significantly in terms of the amount of shift of opinion, the amount of progress toward consensus, and the amount of improvement in quality of judgment effected by discussion.

Sixty subjects, 20 from each of three sections of the course, were organized into 36 groups of five, each group meeting on one or another of three successive days. Discussion was addressed to 39 problems divided into three series of 13 each, one series being discussed during a class period and subjects devoting approximately one minute to reading and two minutes to discussing each problem. Eighteen groups were provided with a moderator instructed to employ full moderation, a total of six moderators being employed; 18 groups had no designated moderator.

Shift in opinion. As each of the 60 subjects recorded a pre- and post-discussion choice of solution on each of 13 problems daily for three days, 2340 opportunities to shift in opinion occurred during the experiment, 1170 in unmoderated and 1170 in moderated discussions. In the unmoderated discussions 375 shifts occurred; under full moderation 309 occurred. Computation of Chi Square yields a value of 8.83, significant at less than the one per cent level of confidence; thus the data indicate that more shift occurred without a designated moderator than in fully moderated discussion.

Consensus. Since each of 36 groups discussed 13 problems, we may speak of 468 discussions, 234 under each experimental condition. For each, pre- and post-discussion consensus scores were computed. The means of the pre-discus-

sion scores under the two experimental conditions (.45 and .43) did not differ significantly. For post-discussion scores the mean under full moderation was .78; that in unmoderated discussion was .86. The *t* test yields a value of 3.48, significant at less than the one per cent level of confidence, and the conclusion may be drawn that more progress toward consensus was made without a designated moderator than in fully moderated discussion.

Judgment. For each of the three hours of participation in the experimental discussions each subject turned in a pre- and a post-discussion score of soundness of judgment, the scores ranging from zero to 13, according as the subject was more or less successful in choosing the preferred solutions of the 13 problems discussed during the hour. The number of scores under each experimental condition was 90. Before discussion the mean score in moderated groups was 6.39; that in unmoderated groups was 6.27. The *t* value of .46 is not significant. After discussion the means were 8.10 and 8.52 in favor of unmoderated discussion. The *t* value of 1.51 is significant at between the 10 and 20 per cent levels of confidence, suggesting that more improvement in quality of judgment may have occurred without a moderator than in fully moderated discussion.

As the result is inconclusive, another approach to analysis of the data may be made. During the experiment there were 2340 instances of a subject participating in discussion of a problem, 1170 under each experimental condition. In each case he could shift from a non-preferred to the preferred solution or start with the preferred solution and hold to it (both categorized as evidence of good judgment); or he could shift from the preferred to a non-preferred solution, shift from one non-preferred solution to

another, or start with a non-preferred solution and hold to it (all categorized as evidence of poor judgment). We may, therefore, classify the subject's judgment in each instance as good or poor. Under full moderation there were 727 instances of good judgment; without a moderator there were 771 instances. The Chi Square test yields a value of 3.43, which is significant at approximately the seven per cent level of confidence. The data strongly suggest, therefore, that better judgment was displayed without a designated moderator than in fully moderated discussion.

EXPERIMENT 2. STYLE OF MODERATION AND LENGTH OF DISCUSSION

The purposes of the experiment were to determine whether two styles of moderation and three lengths of discussion differ significantly in terms of the amount of shift in opinion, the amount of progress toward consensus, and the amount of improvement in quality of judgment effected by discussion.

Sixty subjects, 20 from each of three sections of the course, were organized into 36 groups of five each. Half of the groups had no designated moderator; the other half were provided with moderators instructed to employ full moderation. Each of the six moderators was assigned to an equal number of discussions of the three different lengths. Discussion was addressed to 24 problems organized in nine series in such a way that three of the series included four problems each, three included eight problems each, and three included twelve problems each. Each series was discussed during a class period. Subjects studied each problem one and a half minutes, while the time permitted for discussion of each was two minutes, four and a half minutes, or nine minutes.

Shift of opinion. As each of 60 subjects recorded a pre- and post-discussion

choice of solution for 24 problems, 1440 opportunities to shift occurred during the experiment, 720 in unmoderated groups and 720 in fully moderated groups. In the former 284 shifts occurred, in the latter 279 shifts. Since the Chi Square value (.046) is not significant, the difference in style of moderation did not affect shift of opinion.

The experiment afforded 240 opportunities to shift in 9-minute discussions, 480 opportunities in 4½-minute discussions, and 720 opportunities in 2-minute discussions. In the three categories the percentages of subjects shifting were respectively 42, 39, and 38. Again the Chi Square value (1.26) is not significant, indicating that a hypothesis of no difference cannot be rejected.

Consensus. Pre- and post-discussion consensus scores for the 288 discussions were computed. As expected, analysis of variance of the pre-discussion scores yielded no significant *F* ratios. Mean post-discussion consensus scores for 9-minute, 4½-minute, and 2-minute discussions were respectively .83, .76, .86 in fully moderated discussion, .93, .85, .83 in unmoderated discussion. Analysis of the variance of post-discussion scores is summarized in Table I.

TABLE I
ANALYSIS OF VARIANCE OF POST-DISCUSSION
CONSENSUS SCORES.

Source of Variation	df	Mean Square	F
Style of Moderation	1	.060	1.03
Length of Discussion	2	.095	1.64
Interaction	2	.145	2.50*
Within groups	282	.058	
Total	287		

*Significant at approximately the 10% level.

As appears from the table, neither style of moderation nor length of discussion significantly affected progress toward consensus, though the data sug-

gest interaction between the two variables.

Judgment. As the number of problems constituting a series for discussion during a class period varied, the scores of individual subjects for the hour's discussion were not comparable. The number of members of the group choosing the preferred solution for any one problem was, therefore, taken as the measure of quality of judgment. Groups were composed of five subjects each; hence the scores ranged from zero to five. According to expectation, analysis of variance revealed no significant differences among pre-discussion mean scores for the various experimental conditions. The post-discussion mean scores for 9-minute, 4½-minute, and 2-minute discussions were respectively 2.29, 3.04, 3.28 in fully moderated discussion and 2.50, 3.15, 3.08 in unmoderated discussion. Analysis of the variance of these scores is presented in Table II.

TABLE II
ANALYSIS OF VARIANCE OF POST-DISCUSSION
SCORES OF JUDGMENT.

Source of Variation	df	Mean Square	F
Style of Moderation	1	0	
Length of Discussion	2	12	2.48*
Interaction	2	1	
Within Groups	282	4.84	
Total	287		

*Significant at between the 5% and 10% levels.

Since the analysis of variance suggests, without substantiating, the conclusion that length of discussion affected the amount of improvement in quality of judgment, the data were analyzed also, as in the preceding experiment, in terms of whether the subjects' responses demonstrated good or poor judgment during discussion. Table III presents a breakdown of these data, the left half of the table presenting the data on

TABLE III
QUALITY OF JUDGMENT DISPLAYED IN DISCUSSION

	Full Moderation	No Designated Moderator	9-minute Discussion	4½-minute Discussion	2-minute Discussion
Good Judgment	443	433	119 (50%)	296 (62%)	461 (64%)
Poor Judgment	227	287	121	184	259
Significance of Difference	$\chi^2 = .23$ $.50 < P < .70$			$\chi^2 = 15.35$ $P < .01$	

style of moderation alone, the right half that on length of discussion alone.

It is clear from the table that while style of moderation did not affect quality of judgment, better judgment was displayed in 4½- and 2-minute discussions than in 9-minute discussions.

EXPERIMENT 3. STYLE OF MODERATION AND SIZE OF GROUP

Two styles of moderation and four sizes of groups were tested for their differential effects on shift in opinion, progress toward consensus, and improvement in quality of judgment. Sixty subjects, 15 from each of four sections of the course, were organized into 32 groups, each group size of three, six, nine, and twelve being equally represented. Half of the groups were provided with moderators instructed to provide partial moderation, the other half with moderators instructed to provide full moderation. Eight moderators were employed. Discussion was of 16 prob-

lems arranged in four series of four problems each, subjects being given two minutes to study and seven minutes to discuss each problem.

Shift in opinion. The experiment afforded 480 opportunities to shift in opinion under full and 480 under partial moderation. The percentages shifting were respectively 35 and 30. The frequencies yield a Chi Square value of 2.52, significant at between the 10 and 20 per cent levels of confidence.

The number of opportunities to shift in groups of different sizes were as follows: in groups of three members, 96; in groups of six, 192; in groups of nine, 288; in groups of twelve, 384. The percentages of shifts actually occurring were respectively 31, 33, 32, and 32. The differences obviously are not significant.

Table IV presents the data on shift in opinion so arranged as to indicate whether style of moderation affected shift in opinion more markedly in groups of one size than of another.

TABLE IV
INFLUENCE OF STYLE OF MODERATION AND SIZE OF GROUP ON SHIFT IN OPINION.

Size of Group	3		6		9		12	
	Full	Partial	Full	Partial	Full	Partial	Full	Partial
Style of Moderation								
Number Shifting	17	13	27	36	52	39	70	54
Number not Shifting	31	35	69	60	92	105	122	138
Significance of Difference	$\chi^2 = .44$ $.50 < P < .70$		$\chi^2 = 1.52$ $.20 < P < .30$		$\chi^2 = 2.31$ $.10 < P < .20$		$\chi^2 = 2.69$ $.10 < P < .20$	

In summary, size of group did not affect the proportion of subjects shifting in opinion during discussion. Style of moderation did not significantly affect the proportion shifting in groups of three and six, but the data suggest that in groups of nine and twelve more shift may have occurred under full than under partial moderation.

Consensus. Mean pre-discussion consensus scores for the six experimental conditions did not differ significantly. Mean post-discussion scores for groups of three, six, nine and twelve under full and partial moderation were respectively .55 and .72; .48 and .62; .63 and .45; .58 and .52. Analysis of variance of the post-discussion scores is summarized in Table V.

TABLE V
ANALYSIS OF VARIANCE OF POST-DISCUSSION
CONSENSUS SCORES.

Source of Variation	df	Mean Square	F
Style of Moderation	1	.01	
Size of Group	3	.06	
Interaction	3	.23	1.35
Within Groups	120	.17	
Total	127		

Neither style of moderation nor size of group significantly affected progress toward consensus. A special study perhaps should be made of the possibility, slightly suggested by the data, that the two variables may interact, partial moderation being more conducive to consensus in groups of three and six and full moderation in groups of nine and twelve.

Judgment. As the size of group varied, the measure of quality of judgment was taken as the individual subject's score of right choices of solution for a series of four problems, the scores thus ranging from zero to four. As expected, no significant differences appeared among the

pre-discussion mean judgment scores for the different experimental conditions. The mean post-discussion judgment scores for groups of three, six, nine, and twelve under full and partial moderation were respectively 1.83 and 2.17; 2.50 and 2.50; 2.89 and 2.50; 3.19 and 2.27. Analysis of variance of the post-discussion scores is summarized in Table VI.

TABLE VI
ANALYSIS OF VARIANCE OF POST-DISCUSSION
SCORES OF JUDGMENT.

Source of Variation	df	Mean Square	F
Style of Moderation	1	12.15	12.27**
Size of Group	3	3.79	3.83*
Interaction	3	3.80	3.84*
Within Groups	232	.99	
Total	239		

*Significant at approximately the 1% level.

**Significant at the .1% level.

Analysis of the data yields several conclusions of interest. First, more improvement in quality of judgment was effected by discussion under full than under partial moderation. Second, the amount of improvement varied directly with the size of the group. Finally, the superiority of full over partial moderation in effecting improvement in quality of judgment varied directly with the size of the group: it was inferior to partial moderation in groups of three members but increasingly superior to partial moderation as the size of the group increased from six to twelve.

EXPERIMENT 4. LENGTH OF DISCUSSION

The purpose of the experiment was to test the effect on shift of opinion, progress toward consensus, and improvement in quality of judgment of three lengths of time permitted for discussion.

Sixty-three subjects, 21 from each of three sections of the course, were organized into 27 groups of seven mem-

bers each, nine groups meeting on each of three successive days. Each group of seven was composed of two subjects who in previous discussion in the course had shown better than average judgment on problems like those to be used in the experiment, two who had shown less than average judgment, and three who had shown about average judgment. Of the last three one was chosen randomly to serve as moderator of his group for the hour. Other members of the group, though acquainted with him as a classmate, had never encountered him before as a moderator. Each of the 27 moderators was used once only. Though instructed to employ full moderation, the moderators were without previous training or experience in moderating discussion. They were permitted not only to vote before and after discussion but also to express their own opinions freely regarding the choice of best solution.

Discussion was addressed to nine problems arranged in three series of three problems each. Both the three series and the nine individual problems were controlled in the design of the experiment. Subjects were given two minutes to study each problem, but the length of time permitted for discussion of each during the hour was varied, one problem being discussed for five minutes, one for nine minutes, and one for 13 minutes. The order in which the three lengths of discussion were encountered during the hour, as well as the three

lengths of discussion themselves, were controlled in the experimental design.

Shift in opinion. As each of the 63 subjects discussed three problems under each experimental condition, 189 opportunities to shift occurred under each condition. Twenty-nine per cent of the subjects shifted during the 5-minute discussions, 33 per cent during the 9-minute discussions, and 32 per cent during the 13-minute discussions. The Chi Square value (.72) reveals no significant differences among these percentages.

Consensus. An analysis of variance of pre-discussion consensus scores, as anticipated, revealed no significant differences. Post-discussion scores were analyzed first by an *F* test and then by *t* tests for comparison of the various pairs of means. The results are tabulated in Table VII.

As appears from the table, the data indicate that 13-minute discussions were more effective than 9-minute discussions and probably more effective than 5-minute discussions in promoting consensus.

Judgment. The data included for each of the 81 experimental discussions a pre-discussion and a post-discussion group score of quality of judgment based on the number of group members choosing the preferred solution. As there were seven subjects in each group, the scores ranged from zero to seven. Analysis of variance of the pre-discussion scores yielded a significant *F* ratio. Analysis of covariance was employed to correct for the influence of these pre-dis-

TABLE VII
COMPARISON OF MEANS OF POST-DISCUSSION CONSENSUS SCORES FOR
PAIRS OF EXPERIMENTAL CONDITIONS.

	5-min. disc.	9-min. disc.	5-min. disc.	13-min. disc.	9-min. disc.	13-min. disc.
Means	.68	.58	.68	.82	.58	.82
Significance of Difference	$t = .95^2$	$.30 < P < .40$	$t = 1.55^6$	$.10 < P < .20$	$t = 2.308$	$.02 < P < .05$

cussion differences on post-discussion scores. The analysis is summarized in Table VIII.

TABLE VIII
ANALYSIS OF COVARIANCE OF POST-DISCUSSION SCORES OF JUDGMENT.

Source of Variation	Sum of Squares of Error of Estimate	df	Mean Square	F
Adjusted Means	24.72	2	12.36	3.62*
Within Groups	262.67	77	3.41	
Total	287.41	79		

*Significant at between the 1% and 5% levels.

Analysis of covariance reveals a significant difference among the adjusted means (5.72 for 5-minute discussions, 4.82 for 9-minute discussions, and 6.20 for 13-minute discussions). Computation of *t* values indicates that 9-minute discussion was less effective than 13-minute discussion ($.01 < P < .02$) and probably less effective than 5-minute discussion ($.05 < P < .10$) in promoting improvement in quality of judgment. The data therefore suggest that the three lengths of discussion, in the descending order of their effectiveness in promoting improvement in judgment, may be arranged as follows: 13 minutes, five minutes, nine minutes. This is identical with the order in which the three time lengths promoted progress toward consensus.

SUMMARY AND DISCUSSION

Style of moderation. The three styles of moderation under investigation in these studies may perhaps be regarded as points on a continuum ranging from no moderation through partial to full moderation. The conception must be employed with caution as the no-designated-moderator condition does not preclude the possibility of a member of the group emerging spontaneously to

perform a moderator's functions. But held tentatively, the view affords a possible explanation of some of the findings.

In 2-minute discussions by groups of 5 each (Experiment 1) full moderation apparently obstructed group process: more shift in opinion, more progress toward consensus, and probably more improvement in quality of judgment occurred without a designated moderator than under full moderation. We may conjecture that in so brief a period of time as two minutes group members can do little more than indicate their choice of solution; there is no opportunity for the detailed presentation and examination of argument. It may well be that under these conditions majority influence rather than discussion is the important factor in group decision. This influence brings minority members into line, thus increasing the degree of consensus; and as there is evidence that the majority is somewhat more likely to be right than wrong on this type of problem, majority influence may also improve quality of judgment. The exertion of such influence would be hindered by a moderator's attempt to intervene, especially as he would be likely to insure for minority members more opportunity to speak than they might otherwise have. In the longer discussions of Experiment 2 (nine, four and a half, and two minutes, or an average of about five minutes), the apparent superiority of the no-designated-moderator condition disappears, the two styles of moderation making no significant difference in the amount of shift in opinion, the amount of progress toward consensus, or the degree of improvement in quality of judgment.

When we turn to the 7-minute discussions in Experiment 3, in which partial and full moderation were employed

in groups of four different sizes (three, six, nine, and 12 members each), we note a marked tendency for one style of moderation to excel the other, which was superior depending on the size of the group. In small groups (three or six members each) style of moderation made no difference in the amount of shift in opinion, whereas in large groups (nine or 12 members each) full moderation probably effected more shift. The data suggest that more progress toward consensus was made in small groups under partial moderation, while in large groups more progress was made under full moderation. In groups of three more improvement in quality of judgment occurred under partial moderation; in groups of six, nine, and 12 more improvement occurred under full moderation, and the superiority of full over partial moderation was more apparent as the size of the group increased. When data from the four group sizes are pooled, it is evident that significantly more improvement in quality of judgment was effected by full than by partial moderation. The data suggest, therefore, that when the size of the group has passed a certain critical point, not here accurately determined, full moderation is increasingly superior to partial in facilitating group process.

Size of group. When style of moderation is ignored in Experiment 3, we must conclude that size of group did not affect either the proportion of members shifting in opinion or the amount of progress toward consensus. But the amount of improvement in quality of judgment varied directly with the size of the group. It probably may be assumed that when group size has passed a critical point, the value of further increase in size declines, but nothing in the data indicates what that point may be.

Length of discussion. In the groups of five in Experiment 2 different lengths of discussion (nine, four and a half, and two minutes) did not affect the amount either of shift in opinion or progress toward consensus, but better quality of judgment was displayed in 4½- and 2-minute discussions than in 9-minute discussions. In the groups of seven employed in Experiment 4 under full moderation by untrained moderators length of discussion did not affect amount of shift in opinion, but 13-minute was more effective than 9-minute discussion in promoting both progress toward consensus and improvement in quality of judgment, and the data suggest that in these two respects 9-minute discussion was inferior even to 5-minute discussion. Both Experiments 2 and 4 suggest, therefore, that nine minutes may be inferior to either a somewhat shorter or a somewhat longer period of time for discussion of the type of problem employed in these studies. A possible explanation is that in discussion of five minutes or less there is little time to explore the argument and majority influence is strong; as the majority is somewhat more likely to be right than wrong, its influence not only increases consensus, but results in some improvement in judgment. When nine minutes are available for discussion, many arguments are presented and criticized, majority influence wanes, thinking becomes confused, and the quality of judgment declines. If still another four minutes are available, confusion is dissipated, the better argument wins over the poorer, and the quality of judgment reaches its maximum point. Presumably here also there is a critical point beyond which further prolongation of discussion does not improve judgment, but nothing in the evidence suggests what that point may be.

EFFECTS OF CERTAIN TECHNIQUES OF CREDIBILITY UPON AUDIENCE ATTITUDE*

THOMAS S. LUDLUM
Capital University

I. PURPOSE

IT was the purpose of this study to evaluate, experimentally, the effects of certain techniques for increasing the credibility of an argumentative speech to an audience. Credibility was evaluated in terms of effects on the attitudes held by the audience toward the subject under discussion in the speech.

The techniques selected for study can be designated in Aristotelian terms as artistic ethical proofs (elements within the context of the speech itself which are subject to manipulation by the speaker). Five of these techniques were selected on the basis of scores which they received on an item rating scale administered to a group of speech teachers and a group of students of speech.¹

The techniques thus chosen were: (1) acknowledging opposing arguments, (2) manifesting a high degree of integrity, (3) seeming to lead the thoughts of the listeners easily rather than forcing them arbitrarily, (4) making sure alleged facts are consistent with known facts, and accompanying such facts as the listeners might regard with doubt, with materials which reveal their reliability, and (5) showing that the message is as recent and up-to-date as possible. These

techniques were chosen because they received high scores as determinant factors of credibility and because they could be incorporated easily into an argumentative speech.

II. THE SPEECHES

The speeches used in this experiment concerned the 1956 general election, then some eight months away. These speeches presented contentions dealing with three major issues. These issues were selected on the basis of both anticipated importance and availability of specific developmental materials on both sides, Republican and Democrat. Working with materials furnished by the National Headquarters of the Republican and Democratic Parties the following questions were selected for use as issues: (1) Are the Republican agricultural policies preferable to the Democratic agricultural policies? (2) Are the Republican school policies preferable to those of the Democrats? (3) Are the Republican fiscal policies preferable to those of the Democrats?

Four straight-forward argumentative speeches were prepared, two of which were pro-Republican with one limited to support of party policy while the other concentrated on attacking the policies of the Democratic party. The other two speeches were pro-Democratic, again with one supporting the party and the other attacking the opposition policies.

Four speeches in parallel form to the straight-forward argumentative speeches were prepared, the five techniques of

*Based upon a doctoral dissertation completed in 1956 at The Ohio State University. Professor Franklin H. Knower directed the study.

¹For a discussion of this procedure the reader is referred to the dissertation from which this article is abstracted, A Study for Influencing the Credibility of a Communication, available from University Microfilms, Ann Arbor, Michigan (Mic 57-1846).

credibility being incorporated in the following manner.

1. Opposing arguments were acknowledged, using such statements as: "The Democrats contend that . . . this makes sense, but on the other hand we Republicans believe. . . ." This involved using a brief summary in each speech of the arguments used in the opposition speeches.

2. An attempt was made to manifest high integrity by the use of such phrases as: "I won't try to hide the fact that I think the . . . party has the better program. This is why I'm speaking for them. But I can't deny that the other party has done this." Another example was: "I want to deal today only with those facts which my personal investigation has revealed to be solid." The promise, "I'll never try to pull the wool over your eyes," was used twice.

3. In an effort to lead the thoughts of the listeners easily, emphasis was placed on the fact that the issues raised were controversial. The "this or nothing" statements often used in the original speeches were omitted. The "briefing type transitions" (e.g. This is true because . . .) were omitted. Finally the "debate type" conclusions (e.g. Thus we see that because . . .) were omitted.

4. An attempt was made to insure that the facts seemed reasonable. Typical phrases added were: (1) "Let's confine our discussion to those facts which are reasonable," and (2) "Now both parties have made some extreme statements, let's omit these and deal with the facts." In addition to changing the wording, an effort was made to omit those "facts" which seemed less likely to be accepted.

5. An effort was made to show that the message was recent and up-to-date. Thus, the statistics and examples were

dated and such phrases as "the latest information available," were used.

The four speeches and their parallel forms had many identical factors. The same introduction and conclusion were used in all. In addition the problem statements, farm, school and fiscal were exactly the same. The arguments proper contained the only variation. It was here that the differences between attack and support, straight-forward argumentation and the parallel forms, and Republican and Democratic approaches occurred. While each solution argument was unique, consistency with respect to length was maintained.

The speeches were examined by four teachers in the Department of Speech at The Ohio State University and judged to be valid for the purposes of identifying the types represented in this experiment.

III. ATTITUDE SCALE

A Likert-type attitude scale was prepared to measure audience responses to the speeches.²

A split-half correlation of scores on the attitude scale yielded a coefficient of +.79, which when corrected by the Spearman-Brown formula became +.88. The obtained coefficient of correlation was interpreted as indicating relatively high reliability. Later correlations between scores on the pre- and the post-test for the control group of 100 subjects who took the tests one week apart tended to confirm this interpretation. The coefficient of correlation thus obtained was +.78. The instrument was judged to possess face validity since the items came from the stimulus speeches.

² Rensis Likert, "A Technique for the Measurement of Attitudes," *Archives of Psychology*, 140 (1932), 1-55.

IV. EXPERIMENTAL SUBJECTS AND PROCEDURE

All of the subjects used in the experimental phase of this study were college students enrolled in courses in speech. Most were freshmen and sophomores with some junior and seniors and a few graduate students. Of the experimental subjects, 348 were students at The Ohio State University enrolled in an elementary course in speech, while 372 were students enrolled in speech courses at Ohio University. One hundred of the students from Ohio University were selected as control subjects. Thus, this report contains data on 620 subjects serving in experimental groups and 100 additional subjects serving as controls. Teachers at The Ohio State University and Ohio University administered the attitude scale to these subjects.

One week later the experimental subjects heard one of the four speeches. Each subject was asked to record his name and political preference on an information sheet. The name was used only to provide a key for comparison between the pre- and post-test. The subject's political preference was vital in order to provide information on the basic political variable. Information was then presented concerning the speech. At The Ohio State University the subjects were told that they would hear a political speech by a student who was president of the Young Republicans or Young Democrats Club at Ohio University, depending on which speech they heard. At Ohio University the subjects were told that they would hear a political speech by a graduate student at Ohio University who had been president of the Young Republicans or the Young Democrats when he was an undergraduate at another university. After the information was presented, each group listened to one speech. As soon as the

speech had been completed the attitude scales were administered.

V. ANALYSIS OF DATA AND RESULTS

Two statistical instruments were utilized in analyzing the data. These were the *t*-test of significance, of difference of means, and the standard error of measurement.

1. The characteristics of the group as a whole were determined by the compilation of pre-test and post-test scores. The scores for the pre-test and the post-test were each normally distributed, not bi-modally as one might expect.

There was a very definite similarity between the distribution on the pre-test and on the post-test for the whole sample. The theoretical range for the scale was 40 to 200 with a theoretical mean of 120. On the pre-test the actual range was from 40 to 189 with a mean of 124.9 and a standard deviation of 24.1. On the post-test the range was the same as the theoretical range, 40 to 200. The mean was 124.4 and the standard deviation was 21.0.

2. The *t*-test of the significance of a difference between means was used first to determine whether the control and experimental groups were matched. Second, it was used to determine the significance of the shifts in attitudes associated with the variables studied.

The *t*-ratio between the pre-test scores of the control and experimental groups was only .016, obviously not significant. This was construed as indicating that the two groups were closely matched.

In determining the significance of the difference in shift associated with the variables studied, the technique employed involved calculating the amount of shift for each individual. For the experimental group a movement in the direction of the stimulus speech was

considered positive, a movement in the contrary direction, negative. For the control group a change up the scale toward the Republican end was considered a positive shift for those who had scored above the theoretical mean, 120 on the pre-test, and a negative shift for those who had scored below the theoretical mean. Conversely, a change down the scale was considered a negative shift for those who had scored above the theoretical mean and a positive shift for those who had scored below the mean.

There was a decided difference in the shift occurring between the experimental and control groups. The mean shift for the control group was -2.2 points as contrasted with the mean shift for the experimental group of +3.1 points. The *t*-ratio for the difference was 4.61, significant at the 1 per cent level of confidence. Thus, there was a significant difference in the shift made by the experimental group as contrasted with the control group.

The difference in shift between those who heard the straight-forward argumentative speeches and those who heard the parallel forms of these speeches were significant at the 5 per cent level of confidence. The mean shift for those who heard the straight-forward argumentative speeches was +2.9 points. The mean shift for those who heard the speeches incorporating credibility techniques was only +.1 point. The *t*-score between those groups was 2.37 which was significant at the 5 per cent level of confidence with 618 degrees of freedom. It must be noted that the straight-forward argumentative speeches were more effective than the parallel forms. It is also interesting to note that for listeners of the opposing party the argumentative speeches caused less negative shift, -.3 point to -2.7 points

However, the *t*-ratio of 1.31 was not significant at the 5 per cent level of confidence with 216 degrees of freedom.

The most dramatic difference was between the listeners who heard speeches representing the party with which they considered themselves affiliated and the listeners who heard speeches representing the opposing party. The mean shift for the former was +4.9 points, for the latter, -1.4 points. The *t*-ratio between these two group shifts was 4.92, the largest *t* obtained and significant at the 1 per cent level of confidence with 424 degrees of freedom.

Moreover, a comparison of the shift obtained for these two groups with that obtained for the group which declared no party affiliation brought to light some further significant differences. The mean shift for the group who heard their own party's speeches was +4.9 points. The mean shift for the undeclared group listening to these same speeches was +.6 of a point. The *t*-ratio between these two group-shifts was 3.44 which was significant at the 1 per cent level of confidence with 394 degrees of freedom. Thus, the group listening to their own party's speeches not only shifted significantly more than those listening to opposition party speeches but also significantly more than the undeclared middle group.

The mean shift reported above for the opposition group was -1.4 points. The comparison between this shift and the +.6 shift for the undeclared group resulted in a *t*-ratio of 2.02 which was significant at the 5 per cent level of confidence with 404 degrees of freedom.

The difference between the amount of shift for those who heard attack and those who heard support speeches was not statistically significant. The overall effect of the Republican speeches on mixed audiences was not significantly

different from the over-all effect of the Democratic speeches. The mean shifts of Republican and Democratic groups were not significantly different; nor were the shifts of the group who declared their party affiliation and the undeclared group. Table I presents the data which indicated significant differences in shift of attitude. Table II presents the data on differences in shift which were not statistically significant.

vice by which one may "determine by just how much an individual's score on the test given after hearing the argument must differ from his score on the first test to amount to a significant difference."

After the standard error of measurement was calculated, the computation of shifts in attitude was simplified by calculation of the significance of shifts of ten, twenty and thirty points re-

TABLE I
GROUP SHIFTS WHICH WERE SIGNIFICANT.

Group	Number of Subjects	Mean Shift	t ratio	df	Level of Significance
Experimental group.	620	+3.1	4.61	718	1%
Control group.	100	-2.2			
Heard straight-forward argumentative speech.	343	+2.9	2.37	618	5%
Heard argumentative speech revised to utilize credibility techniques.	277	+0.1			
Heard speech representing own party.	208	+4.9	4.92	424	1%
Heard speech representing opposing party.	218	-1.4			
Declared group hearing speech representing own party.	208	+4.9	3.44	394	1%
Undeclared group.	188	+ .6			
Declared group hearing speech representing opposing party.	218	-1.4	2.02	404	5%
Undeclared group.	188	+ .6			

TABLE II
GROUP SHIFTS WHICH WERE NOT SIGNIFICANT AT THE 5 PER CENT LEVEL OF CONFIDENCE.

Group	Number of Subjects	Mean Shift	t ratio	df
Heard speeches attacking opposition policies.	319	+2.0		
Heard speeches supporting party policies.	301	+ .7	1.18	618
Heard Republican speeches.	319	+1.9		
Heard Democratic speeches.	314	+ .6	1.3	618
Republicans	270	+1.4		
Democrats	161	+1.5	.29	429
Declared party affiliation.	431	+1.6		
Did not declare party affiliation.	188	+ .6	.91	617

3. The standard error of measurement was used to analyze individual shifts in attitude. In his report of the effects of oral argument on changes in attitude, Knower recommended this device as deserving the serious consideration of all students of attitude measurement.³ He pointed out that it is a de-

spectively. In terms of the critical ratios of the standard error of measurement a shift of 30 points equaled $2.56\sigma_{\text{meas}}$ and was significant at the 1 per cent level of confidence; a shift of 20 points equaled $1.77\sigma_{\text{meas}}$ and was significant at the 4 per cent level of confidence; and a shift of 10 points equaled $.89\sigma_{\text{meas}}$ and was significant only at the 18 per cent level of confidence.

³ Franklin H. Knower, "A Study of the Effect of Oral Argument on Changes of Attitude," *Journal of Social Psychology*, VI (1935), 327.

Evaluated on this basis, the pattern of individual change in this study was strikingly different from that found in the Knower study, referred to above. In this study only 17 per cent of the experimental subjects shifted positively at least ten points. Only 2.1 per cent shifted positively at least thirty points. There was even less individual shift among control subjects. None of them shifted positively thirty points while 9 per cent shifted positively as much as ten points. Ten per cent of the controls and 9.7 per cent of the experimental subjects shifted at least ten points negatively.

Fairly large differences in individual shift occurred between those who heard their own party's speeches and those who heard speeches of the opposing party. Of the former, 3.9 per cent moved positively thirty points or more. Only 1.4 per cent of the latter group moved this far. Eleven per cent of the favorable audience moved twenty positive points or more while only 4.1 per cent of the opposition audience did. Thirty per cent of the favorable audience shifted more than ten points positively while 19.6 per cent of the opposed did.

The negative shift was more dramatic. While 3.9 per cent of the favorable

group shifted negatively ten points or more, perhaps due to regression toward the mean; 16 per cent of the other group shifted negatively. This later figure is more than four times the percentage of the former and operated in contradiction to the tendency to regress toward the mean. A more complete tabulation of the percentages shifting may be found in Table III.

VI. CONCLUSIONS

1. The results of this study again demonstrated that a significant change in attitudes can be brought about by exposing subjects to a brief oral argument.

2. However, the value of formal political speaking in winning converts to the speaker's party from the opposing party may be overemphasized. The results of the study indicate that the political speaker is not likely to cause those auditors who consider themselves to be affiliated with the opposing party to shift their attitudes to a position more favorable to his party by the presentation of one speech. Considered either as individuals or as a group their shift was slight and often negative.

3. Moreover, the political speaker may well find that those techniques commonly recognized as contributing to

TABLE III
INDIVIDUAL SHIFT DATA FOR EXPERIMENTAL VARIABLES.

Group	Per Cent Shifting				
	Positive		Negative		
	30 pts.	20 pts.	10 pts.	10 pts.	30 pts.
Experimental Group	2.1	6.	17.	9.7	1.3
Control Group	0.	2.	9.	10.	1.
Heard Speech Representing Own Party	3.9	11.	30.3	3.9	0.
Heard Speech Representing Opposing Party	1.4	4.1	19.6	16.0	2.3
Heard straight argumentative speech	2.3	6.9	18.8	2.6	.87
Heard argumentative speech revised to utilize credibility techniques	1.8	5.4	16.	12.9	1.40
Heard attacking speech	1.9	8.1	20.2	10.	.63
Heard supporting speech	2.	3.9	14.8	10.5	1.60
Heard Republican speeches	1.5	5.6	18.	7.6	.9
Heard Democratic speeches	1.9	6.	16.6	11.5	1.9
Declared party preference	2.6	7.5	19.5	11.6	1.4
Did not declare party preference	1.3	2.7	11.7	5.5	1.1

credibility for general communication have little real value for him. A straightforward argumentative approach seems, from the results of this study, to be more effective in political speaking.

4. Furthermore, the political adage that the party in-office should "point

with pride" while the party out-of-office "views with alarm" may not be valid advice. The results would seem to indicate that there is no real difference in effectiveness between speeches of attack on, and speeches in support of party policy.

T

lation
ing
of
lect
spea
exc
per
stu
lati
effe
hav
pos

T
pro
nee
desc
beh
effe
"pu
the
vey
tive
tion
gen
rese

•
vers
by
¹
eral
²
desc
con
una
abil
³
"Th
⁴
Gu

THE RELATIONSHIP BETWEEN THE ABILITY TO SPEAK EFFECTIVELY AND THE PRIMARY MENTAL ABILITIES, VERBAL COMPREHENSION AND GENERAL REASONING*

JOE M. BALL

University of Pittsburgh

INTRODUCTION

THE literature in the field of speech is replete with studies of the relationship between measures of speaking effectiveness and various dimensions of the physical, emotional, and intellectual behavior of persons doing the speaking. Gilkinson¹ has published an excellent summary of the literature pertinent to these areas. Representative studies seeking to determine the relationship between a measure of speech effectiveness and general intelligence have indicated consistently that a low, positive correlation exists.

There were two considerations which prompted the present study: (1) the need for a more precise and definitive description of the role which intellectual behavior plays in the ability to speak effectively and (2) the availability of "pure"² factor tests, as represented by the Guilford-Zimmerman Aptitude Survey,³ which appear to offer more sensitive instrumentation for the exploration of these relationships than did the general intelligence tests used in earlier research.⁴

* Based upon Ph.D. dissertation, The University of Southern California, 1951, directed by Lee Edward Travis.

¹ H. Gilkinson, *Outlines of Research in General Speech*, Minneapolis, 1943.

² The use of the term "pure factor" in describing a test indicates that the test is so constructed and designed that, within certain unavoidable limits, it measures one factor or ability and one only.

³ J. P. Guilford and Wayne S. Zimmerman, "The Guilford-Zimmerman Aptitude Survey."

⁴ See also Guilford and Zimmerman, "The Guilford-Zimmerman Aptitude Survey," *Journal*

This study has limited itself to ascertaining the relationship between speaking ability and two of the primary mental abilities, Verbal Comprehension and General Reasoning.

PROCEDURES

Measuring Instruments. For the purposes of this study, speaking ability has been defined as the rated ability of an individual to explain, clarify, or prove the point of a one point speech.

In order to measure the ability of an individual in Verbal Comprehension and General Reasoning, parts I and II of the Guilford-Zimmerman Aptitude Survey were administered.

This battery was selected with an eye for the "pure factor" nature of these sub-tests, a selection which was justified by the low order of their intercorrelation. For the subjects of this study, the r values were for the males .29, for the females .12, and for the combined sexes .23. These intercorrelations indicate that a subject did not need a high level of aptitude for one test in order to attain a high score on the other.

The Sample. All students enrolled in the beginning public speaking course at the University of Southern California during the Spring Semester of 1948 were used as subjects in this study. The total number of subjects was 221 (187 males and 34 females).

Ages for males and females combined

of Applied Psychology, 32:24-35, February, 1948, for discussion of the advantages in using "pure factor" tests over general intelligence tests.

TABLE I
MEANS AND STANDARD DEVIATIONS FOR THE AGES AND NUMBER OF YEARS OF SCHOOLING
OF MALES AND FEMALES, SEPARATELY AND COMBINED.

	Males (187)		Females (34)		Combined (221)	
	M	σ	M	σ	M	σ
Ages	23.6	1.3	20.2	.4	24.1	1.1
Years of Schooling	14.2	.4	13.8	.1	13.7	.4

ranged from eighteen to forty-five years, with a mean of 24.1, and a standard deviation of 1.12. The number of years of schooling ranged from nine to nineteen, with a mean of 13.7 and a standard deviation of .41. Table I presents the distribution of ages, and the years of schooling for the subjects used in this study.

Measurement of Verbal Comprehension and General Reasoning. Parts I and II of the Guilford-Zimmerman Aptitude Survey were used to measure the aptitude of each subject for verbal comprehension (Part I) and general reasoning (Part II).

Prior to taking the tests, the subjects were told that data gathered would be used by their instructors to better evaluate individual and group problems in connection with class work.

Judging the Speaking Ability of Each Subject. The speaking effectiveness of each subject was determined on a basis of the cumulative judgments of his classmates.⁵

The speech used in the judgment situation was a three-to-five minute talk developing one point through the use of verbal forms of support. The emphasis in this speech was placed on the use of statistics clarified by the analogy. This

pattern was a standard class assignment in all twelve sections. Each subject had the choice of developing the main point of his speech along either inductive or deductive lines.

The judgments of speaking effectiveness were expressed in terms of a whole number selected from a seven-point rating scale. The scale ranged from 1—Among most effective to 7—Among least effective. The judges recorded their evaluations of each speaker on a separate ballot. The subject's rating as a speaker was the mean of all judgments rendered. Prior to the ratings of the speakers, each instructor made it clear to his section that the judgments rendered on a speaker would have no influence on the grades assigned by the instructor to each speaker. Figure 1 is a reproduction of the rating scale employed in this study.

In order to estimate the reliability of the speech ratings a coefficient of correlation was computed between odd and even speech ratings received by all subjects. The first twelve ratings received by any subject were selected as the basis for comparison. The resulting coefficient of correlation was .79, corrected by the Spearman-Brown formula to .82. An r of this size for a population of two hundred and twenty-one indicates a fairly high level of agreement among the student judges.

ANALYSIS OF DATA

I. *Characteristics of Group Scores on the Measured and Rated Variables.* Table II presents the means, standard

⁵ For discussion of the use of students in rating general effectiveness of speaking, see:

A. H. Monroe, H. H. Remmers, and E. V. Lyle, "Measuring the Effectiveness of Public Speech in a Beginning Course," *Bulletin of Purdue University*, 37, *Studies in Higher Education*, 29:27, September, 1936.

A. I. Bryan and W. H. Wilkes, "Audience Tendencies in Rating Public Speakers," *Journal of Applied Psychology*, 26:371-381, June, 1942.

FIGURE I

Name of Speaker	Section	Date	Rating			
(1) Among most effective	(2) Among more effective	(3) Among fairly effective	(4) Those midway between	(5) Among barely effective	(6) Among less effective	(7) Among least effective

When the speaker has finished his speech, judge him immediately as to how well he EXPLAINED, PROVED, OR CLARIFIED HIS MAIN POINT: mark your rating with a whole number according to the scale above and place the rating in the upper right-hand corner.

TABLE II

MEANS, STANDARD DEVIATIONS AND STANDARD ERRORS OF THE MEANS OF RAW SCORES ON THE VERBAL COMPREHENSION AND GENERAL REASONING APTITUDE TESTS AND SPEECH RATINGS RECEIVED FOR MALES, FEMALES, AND BOTH SEXES COMBINED.

Aptitude Test	VCx			GRxx			SRRxxx		
	M	σ	σM	M	σ	σM	M	σ	σM
Male (N-187)	35.18	13.48	.98	11.58	4.67	.34	2.84	.83	.06
Female (N-34)	35.02	12.12	2.11	8.14	5.57	.98	2.86	.82	.15
Combined (N-221)	35.80	12.50	.85	11.02	4.70	.32	2.84	.84	.06

xVC—verbal comprehension

xxGR—general reasoning

xxxSRR—speech ratings received

deviations and standard errors of the means of raw scores on the aptitude tests and speech ratings received for males, females, and the combined sexes.

Scores on the Verbal Comprehension and General Reasoning Tests were determined by the scoring formula included in published directions for the administration of the Guilford-Zimmerman Aptitude Survey.

Speech ratings for each subject were computed by taking the average of all speech ratings the subject received. The number of such ratings varied from subject to subject in accordance with the size of his particular section and the number of absences from his section during the rating sessions. The smallest number of ratings received by any subject was twelve, the greatest number was twenty one, the average was fifteen. Examination of Table II reveals a tendency for the distribution of speech ratings to cluster towards the superior end of the scale. Student judges in all sections

tended to ignore inferior ratings of six and seven. In essence this resulted in a five-point rating scale instead of seven points with the mid-point of three instead of at four.

II. Analysis of Inter-Section and Inter-Sex Differences on the Variables. Since the subjects for this study were drawn from twelve public speaking classes and included female as well as male subjects, it was necessary to evaluate the possibility of inter-section and inter-sex differences on the variables before any comparison among the variables could be made.

The analysis of variance technique described by McNemar⁶ was used to evaluate inter-section differences. The results of this analysis indicated that there were no sectional differences on the aptitude tests scores nor on speech ratings.

⁶ McNemar, Q., *Psychological Statistics*, New York: John Wiley and Sons, 1949, p. 236.

In order to determine possible male and female differences on the variables, t-ratios were computed between the differences of the means of the sexes for each variable. The only significant difference between the sexes was found on the General Reasoning test scores.

Because of the difference between the sexes on the General Reasoning test scores and because of other differences revealed by statistical analysis which lies outside the scope of this report, the raw scores on all rated or measured variables were converted into a standard score scale. The stanine scale was selected because of its simplicity and adequacy of spread.

III. Comparison between the aptitude test scores of the upper and lower twenty-seven per cent of speech ratings received by the males. The upper twenty-seven per cent of males on the speech ratings were found among stanines nine, eight, and seven. These stanines represented those males whose speaking ability was rated among those more or most effective. The lower twenty-seven per cent were found in stanines three, two, and one or those males rated among those less or least effective in speaking. The means of scores on the aptitude tests for both groups were then compared by computing t-ratios between the differences of the means.

The t-ratio for Verbal Comprehension was found to be 3.24 and for General Reasoning 3.62. With one hundred subjects as the basis for comparison, a t-ratio to be significant at the .01 level had to exceed 2.63.⁷

IV. Coefficients of Correlation between Speech Ratings and the Aptitude Test Scores. The stanine values were used to

compute Pearson product-moment coefficients of correlation among the variables. Inspection of Table III reveals that for the male subjects of this study the intercorrelations of .25 and .24 respectively indicate that a small but significant relationship exists between Speech Ratings and both Verbal Comprehension and General Reasoning. These correlations were significant at the .01 level of confidence.

Further inspection of Table III shows that there was practically an absence of correlation between Speech Ratings for the females and the aptitude test scores. This difference in correlation for the males as opposed to the females could be due to the small number of females in each section and the overall small sample of females. It might be that female speakers in being received well are dependent upon different factors or skills from the males.

V. Predicting speech ratings when verbal comprehension and general reasoning tests scores are known. Predicting the speech rating an individual would receive knowing his Verbal Comprehension and General Reasoning test scores required the computation of a multiple coefficient of correlation, R, between the speech ratings and a combination of the scores on the aptitude variables with their inter-correlations considered. The multiple correlation technique described by Guilford⁸ was used in this computation.

Table IV presents the results of this computation. Multiple coefficients of correlation, R, are subject to the same interpretation as the coefficient of correlation, r. Inspection of Table IV reveals that for the males and both sexes combined the R of .31 was significant at the .01 level of confidence, but the extent of the relationship would give

⁷ Guilford, J. P., *Fundamental Statistics in Psychology and Education*, New York: McGraw-Hill, 1956, pp. 538-39.

⁸ Guilford, *op. cit.*, pp. 401-404.

TABLE III

COEFFICIENTS OF CORRELATION AND STANDARD ERRORS OF CORRELATION BETWEEN SPEECH RATINGS AND THE APTITUDE TEST SCORES FOR MALES, FEMALES, AND THE COMBINED SEXES.

Aptitude Tests	Speech Ratings		
	Males	Females	Combined
VC*	.25 ± .06	.02	.21 ± .06
GR**	.24 ± .06	.02	.21 ± .06

* Verbal Comprehension.

** General Reasoning.

r's needed for .01 level of significance

Males (N=187)	.19
Females (N=34)	.43
Combined Sexes (N=221)	.15

TABLE IV

MULTIPLE CORRELATIONS, STANDARD ERRORS OF CORRELATION, COEFFICIENTS OF DETERMINATION, AND INDICES OF FORECASTING EFFICIENCY BETWEEN SPEECH RATINGS AND A COMBINATION OF THE APTITUDE TESTS SCORES, WITH THEIR INTERCORRELATIONS CONSIDERED, FOR THE MALES, AND THE COMBINED SEXES.

Subjects	r			R	R ²	E
	VC and SRR	VC and GR	GR and SRR	σ_R		
Males (N=187)	.25	.29	.25	.31*	.09	
Combined (N=221)	.22	.23	.21	.27	.07	.05
				.06	.07	.02

* Significant at the .01 level.

little assistance in predicting speech ratings from a combination of the aptitude test scores.

A coefficient of determination, R², was also computed and yielded a value of .09 which indicates that the variance present in the two aptitude tests is accounting for less than ten per cent of the variance present in the speech ratings received by the males. Further, the index of forecasting efficiency, E (.05), when expressed as a percentage shows that our margin of error in predicting speech ratings from the aptitude tests scores is only five per cent less than would be the errors of prediction without the tests.

SUMMARY AND CONCLUSIONS

The purpose of this study was to ascertain the relationship between speak-

ing ability and two of the primary mental abilities, Verbal Comprehension and General Reasoning. The following conclusions were reached as a result of this study:

(1) There is a small but significant relationship for the subjects of this study between their speaking ability and their Verbal and Reasoning abilities. The extent of the relationship is too slight, however, to allow for accurate predictive data. The predictive indices established from the data are only slight improvements over chance predictions.

(2) When considered separately, both Verbal Comprehension and General Reasoning show the same degree of low, positive correlation with speaking ability for the male subjects and a practical absence of correlation for the female subjects. Neither test may be

said to contribute more than the other to speaking ability.

(3) The low, positive correlation between speaking ability and the aptitude tests for the males as opposed to the practical lack of correlation for the females on these variables could have been due to the small number of females per section and the overall small sample of females. It might also indicate that female speakers in being received well are dependent upon different skills or abilities from the males. The former is

the more probable explanation for this difference.

(4) The results of this study, when compared with results obtained in previous investigations of the relationship between speaking ability and general intelligence, indicate a large measure of agreement in that low, positive correlations were the trend. This may be accounted for by the probability that the various measures of general intelligence used in prior investigations contained significant loadings of verbal and reasoning factors in their construction.

ORAL AND NASAL SOUND PRESSURE LEVELS OF VOWELS

RAYMOND SUMMERS
Indiana State Board of Health

THE purpose of this experiment was to compare oral and nasal sound pressures of eight vowel sounds which were produced at four specified sound pressure levels by male and female subjects.

I. SUBJECTS, APPARATUS, AND PROCEDURE

Thirty subjects were used in the experiment, 16 males and 14 females. Their ages ranged from 20 to 41 years. Professional listeners judged the subjects as having typical, General American vowel production and good voice quality (with particular emphasis placed on freedom from nasality). They also possessed an ability to produce vowels at specified sound pressure levels and they demonstrated an ability to prolong isolated speech sounds.

Two independent recording systems, one of which included a probe-tube, were used to record the speech samples. The probe-tube was a three-inch length of seamless nickel tubing which was coupled to the diaphragm of a Western Electric Model 640-AA condenser microphone. The length of the probe-tube was partially determined by the fact that the coupler, condenser microphone, and its associated pre-amplifier (Audio Instrument Company Model 16W) should be placed below the mouth in order to minimize interference with the sound field.

To quantify the oral and nasal sound pressures, graphic recordings of the speech samples were obtained using a

Sound Apparatus Company High Speed Level Recorder (HPL-E). Also used were an Ampex Recorder and a Stancil-Hoffman Recorder, the same instruments used to record the speech samples originally. A paper speed of 10 mm/second was employed on the HPL-E and, using the maximum damping setting, the graphic recording tended to be smoothed.

The vowels used in this experiment were [i/ e/ ε/ æ/ a/ ɔ/ o/ u]. Although [e] and [o] are frequently the diphthongs, [er] and [ou], the subjects attempted to produce the pure vowels. The four sound pressures studied were 60, 70, 80, and 90 decibels (re .0002 dyne/cm²) when the microphone of the sound level meter was eight inches from the subject's mouth.

One important aspect of this experiment involved two independent recording systems which were essentially identical in their frequency response. These two systems were calibrated with random noise in a sound treated room. Since the type of calibration depends upon the use of the probe-tube, a free field calibration of the probe coupled to the microphone was employed.¹

Two sound treated rooms, with a window in the wall between them, were used for recording the vowel sounds. The two recorders and a technician occupied one room and the subject and experimenter occupied the other. The subjects, upon arriving for their individual appointments, sat in a medical

*Based upon Ph.D. dissertation, Purdue University, 1955, directed by M. D. Steer.

¹R. W. Benson, "The Calibration and Use of Probe-tube Microphones," *Jour. of Acoustical Society of America*, XXV (1953), 128-134.

type examination chair in such a position that they could observe a sound level meter which indicated the sound pressure at which the vowels were being produced. The microphones of the sound level meter and the oral recording system were juxtaposed and maintained at a distance of eight inches from the mouth for all subjects. These two microphones were within $1/4$ th of an inch of each other so that both were influenced by approximately the same sound pressure field.²

After a practice period, during which the subjects produced vowels at sound pressure levels of 60 to 90 db, the probe-tube was inserted into the subject's nasal meatus approximately $1/4$ th of an inch.

The vowels to be produced and the sound pressure levels to be attained were written on cards and were visible to the subjects. Both sounds and levels were randomly presented although all vowels within one sound pressure level were produced successively.

In the analysis of the graphic sound pressure levels, the HPL-E graphs were aligned vertically, by time, so that measurements of the same portion of the oral and nasal recordings were made. To measure the same duration for each vowel, the experimenter marked off four-centimeter lengths on the HPL-E graphs representing four seconds of recording. Three factors were considered in locating the section to be measured. The section measured was not at either the initiation or cessation of the recording, constant sound pressure was indicated, and continuous phonation was manifested.

²The "oral" sound pressure, obtained eight inches from the mouth, actually constituted an "overall" sound pressure since it was influenced by sound coming from both the mouth and nose. It will be referred to as the "oral" sound pressure while the sound pressure obtained with the probe-tube will be referred to as being "nasal."

The sound pressure measurements on the HPL-E graphs were made by visual inspection and all were verified by an independent examiner. To check the experimenter's reliability, a third person measured the 64 sound pressure levels of one subject and a coefficient of reliability between the experimenter and this third person was .99.

II. RESULTS

Oral and nasal sound pressures of eight vowels which were produced at four sound pressure levels (SPLs) by 30 subjects were collected. The data were treated by an analysis of variance with subjects, groups, vowels, sound pressure levels, and locations constituting the main effects. All of these were treated as fixed variables except subjects which constituted a random variable. The error term was the interaction of subjects-within-groups with the appropriate main effects and interactions. Table I is the analysis of variance for the total experiment.

Since the experimenter was primarily interested in comparing relative differences in oral and nasal SPLs, the main effect of location and the first-order interactions involving location were of importance.³

Location. Location refers to the nasal meatus (nasal location) and a position eight inches from the mouth (oral location). When averaged across groups, vowels, and SPLs, the mean oral sound pressure level was 71.4 db and the mean nasal sound pressure level was 104.1 db, a difference which is significant at the one per cent level. This difference, partially attributed to the fact that the oral SPLs were obtained eight inches from the mouth whereas nasals SPLs were ob-

³Only first-order interactions are reported here. All possible interactions were studied and they are reported in the dissertation.

TABLE I

ANALYSIS OF VARIANCE FOR THE MAIN EFFORTS OF SUBJECTS (S), GROUPS (G), VOWELS (V), SOUND PRESSURE LEVELS (P), AND LOCATIONS (L), AND ALL POSSIBLE INTERACTIONS.

Source of Variation	df	Mean Square	F*
G	1	16.04	.08
Subjects-within-groups	28	203.99	
V	7	38.85	6.97
V x G	7	6.30	1.13
V x Subjects-within-groups	196	5.57	
P	33	45.198.91	1,230.57
P x G	3	83.99	2.29
P x Subjects-within-groups	84	36.73	
L	1	511.038.01	4,330.46
L x G	1	1,691.77	14.34
L x Subjects-within-groups	28	118.01	
V x P	21	6.81	2.29
V x P x G	21	4.74	1.60
V x P x Subjects-within-groups	588	2.97	
P x L	3	1,240.39	49.85
P x L x G	3	114.05	4.58
P x L x Subjects-within-groups	84	24.88	
V x L	7	54.19	12.46
V x L x G	7	10.30	2.37
V x L x Subjects-within-groups	196	4.35	
V x L x P	21	3.74	1.76
V x L x P x G	21	3.86	1.82
V x L x P x Subjects-within-groups	588	2.12	
Total		1.919	

*The critical points for tests with the following degrees of freedom:

df	1%	5%
1.28	7.64	4.20
7,200	2.73	2.05
3,80	4.04	2.72
20,400	1.92	1.60

tained well within the nasal meatus, is so great that a formal test of differences was unnecessary.⁴

First-Order Interactions Involving Location. All two-factor interactions involving location were significant at the one per cent level. When averaged across vowels and nominal⁵ sound pressure

⁴Due to the large difference between the oral and nasal sound pressures, any term involving locations may violate the assumption of additivity which underlies the analysis of variance.

levels, the mean sound pressure levels for the oral location were 72.2 db and 70.5 db for the males and females respectively and, for the nasal location, 103.1 db and 105.2 db for the same respective groups. Thus, the female subjects had a lower oral sound pressure level and a higher nasal sound pressure level than the males. According to Table II, the mean figures for the SPL-by-location interaction indicate that the differences between the nominal SPLs for the oral recordings are greater than the differences between the SPLs for the nasal recordings. The vowel-by-location interaction (Figure 1) indicates not only that the nasal SPL for each vowel is considerably higher than the oral SPL, but also that when an attempt is made to hold the oral sound pressure at a constant level, some factor is operating which makes the nasal sound pressure

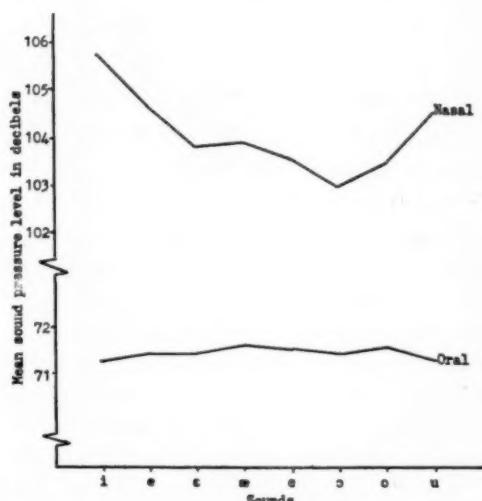


FIG. 1.—Graphic representation of the vowel-by-location interaction (significant at the one per cent level). Oral microphone was located eight inches from the mouth and the nasal, probe-tube microphone was inserted in the nasal meatus.

⁵The actual SPLs were 58.4, 66.9, 76.0, and 84.4 db (re .0002 dyne/cm²) and these are referred to in this study as "nominal" sound pressures of 60, 70, 80, and 90 db respectively.

TABLE II
RECORDED MEAN ORAL AND NASAL SOUND
PRESSURE LEVELS IN DB (RE .0002 DYNE/
CM²) WHEN AVERAGED OVER ALL SOUNDS
FOR EACH NOMINAL SOUND PRESSURE
FOR THE TOTAL GROUP

Location	Nominal Sound Pressure Level	Sound Pressure Level for Total Group
Oral	60	58.4
Oral	70	66.9
Oral	80	76.0
Oral	90	84.4
Nasal	60	93.7
Nasal	70	101.9
Nasal	80	108.0
Nasal	90	112.7

for the high front vowels and the high back vowels greater than that for the middle vowels, a pattern similar to the vowel triangle with the exception that the lowest SPL is shifted to [ɔ] instead of being at [ɑ].⁶ This is graphically represented in Figure 1. This might be attributed to the variation in the cross-sectional area of the mouth during the production of the vowels, i.e., open mouth for low vowels and consequently low pressure and close mouth for high vowels and thus higher pressure.

Groups. Groups consisted of male and female subjects who were randomly selected from a college population. When averaged across vowels, SPLs, and locations, the mean SPL for the females was 87.9 db and for the males, 87.7 db. According to the analysis of variance, the difference between the groups is not significant. With a visual cue to indicate when the proper sound pressure is obtained, one would expect the mean sound pressures to be similar for the two groups.

⁶In interpreting the graphic representations of the interactions, attention should be called to the parallel character or lack of parallel character which is represented. It should be noted that if the lines connecting points on the graphs are parallel from vowel to vowel, the interaction is not significant. If the lines are not parallel on the graphs, the interaction is significant.

Vowels. Eight vowel sounds were studied and when averaged across groups, sound pressure levels, and locations, the mean SPLs for these vowels were: [i], 88.5; [e], 88.0; [ɛ], 87.6; [æ], 87.7; [ɑ], 87.6; [ɔ], 87.1; [o], 87.5; [u], 88.0. The analysis of variance indicates that the differences among these sound pressures are significant at the one per cent level.

Sound Pressure Levels. Each vowel was produced at four specified sound pressure levels. When averaged across groups, vowels, and locations, the mean SPLs were 76.0, 84.4, 92.0, and 98.5 db (re .0002 dyne/cm²) for the 60, 70, 80, and 90 db levels respectively. These differences are significant at the one per cent level.

The discrepancy between the 60, 70, 80, and 90 db nominal sound pressure levels and the exact, mean sound pressure levels of 58.4, 66.9, 76.0, and 84.4 db (re .0002 dyne/cm²) respectively at which the vowels were actually produced is attributed to the use of random noise in the calibration process. It was not expected that the production of vowel sounds at these SPLs would necessarily be equivalent to the random noise signals. Thus, the difference between predetermined SPLs, based on speech signals, and nominal SPLs, based on random noise, is understandable.

Other first-order interactions were studied also. Two of these interactions, vowel-by-group and SPL-by-group, are not significant. The vowels-by-SPL interaction (Figure 2) is significant at the one per cent level.

III. SUMMARY AND CONCLUSIONS

In this experiment, 30 subjects phonated eight different vowel sounds at four specified sound pressure levels and these sounds were recorded by two independent recording systems, a probe-tube

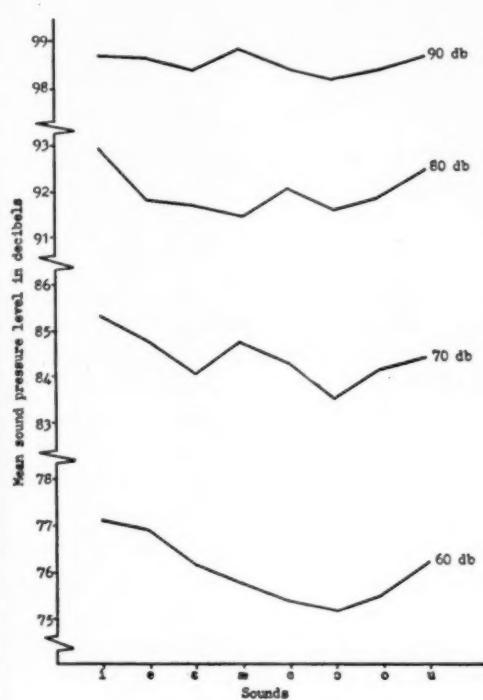


FIG. 2—Graphic representation of the vowel-by-sound pressure level interaction (significant at the one per cent level).

microphone being used to pick up the sound pressure in the nasal meatus and a conventional, studio microphone being used to pick up the sound pressure eight inches from the mouth.

After recording the speech samples, graphs of the sound pressures were recorded on HPL-E tapes and subsequent sound pressures for each subject were obtained. The statistical treatment of the data was by analysis of variance.

The following conclusions can be made:

(1) In regard to the hypotheses which were set out to be tested, there are no

significant sound pressure differences between the two groups studied. However, there are significant sound pressure differences among the eight vowels that were studied, among the four sound pressure levels that were studied, and between the two locations that were studied.

(2) There are differences between and within oral and nasal sound pressure levels when different vowels are produced at different sound pressure levels by male and female subjects.

(3) The differences in sound pressure, within the conditions under which the data were collected, are attributable to differential effects of the primary sources of variation, and to interactions among these variables. A consideration of these main effects and the first-order interactions resulted in the following relationships: (a) The females have a lower oral SPL, across all sounds and SPLs, than the males. The converse is true for the nasal SPLs. (b) The differences among the oral SPLs, for both males and females, are greater than the differences among the nasal SPLs. (c) The mean of the four nasal SPLs for each vowel, when plotted, resembles the conventional vowel triangle except that the lowest sound pressure is for [ɔ]. Probably because of visual monitoring, this relationship does not exist for the oral sound pressure. (d) There are no sound pressure differences between male and female groups in producing vowel sounds at the four nominal SPLs. (e) The mean oral and nasal SPL for each vowel indicates that differences exist among the vowel sounds at the four nominal SPLs studied.

OBSERVED AND PREDICTED ESTIMATES OF RELIABILITY OF ASPECTS OF A SPEECH ARTICULATION RATING SCALE

NANCY MARCH, CARL H. WEAVER,¹ SHEILA MORRISON, and JOHN W. BLACK
The Ohio State University

THE purpose of the study reported here was to determine the economies that might be introduced into a particular speech-rating technique,* economies in judges, items, and subjects. Procedures relevant to this objective provide an opportunity to illustrate and compare some of the alternative methods that are available to the researcher who is confronted with matters of reliability in tests that involve ratings by a panel of judges.

THE SCORES

Morrison employed a rating procedure, modeled upon the method of equal-appearing intervals, to obtain scaled samples of children's articulation. The ratings ranged from *one* (good) to *nine* (poor) in integer units.² This development was patterned after Thurstone's work in scaling attitudes³ and exploited the work of Lewis and Sherman who scaled samples of stuttering in terms of severity.⁴ Morrison's graded samples of children's speech have been used in training speech therapists to assess children's articulation and—of more relevance to the present report—in researches that require the training

of judges to evaluate the merit of articulation in children's speech.

In the present instance the samples that were evaluated were from conversational speech that was elicited from forty children before and after a term of speech therapy. For the most part, the therapy extended over a ten-week term with a single three-hour session of therapy each week. The recorded material was edited to remove the speech of the interviewer, and the remaining segments of speech were adjoined to make one minute of continuous speech per subject. This record was then segmented into six ten-second samples (items). The 480 items (forty subjects, six items per subject for both pre- and post-therapy conditions) were randomized for presentation to ten judges, who were advanced students in speech and hearing.

The judges were trained in a manner recommended by Morrison. A training session of about one and one-half hours occurred on the day preceding the experimental judging session. Judges listened to samples of children's speech that illustrated the range of the scale, and were then given instructions for rating similar segments of speech: that the rating should be based on articulation throughout the segment; that the judges should avoid the influence of general effectiveness insofar as possible; and that the scale extended from *one*, representing the least defective articulation, to *nine*, the most defective articulation. The judges then made practice judgments of thirty-six samples that were arranged in random order. These

*Part of this study derives directly from a master's thesis by Nancy March under the direction of John W. Black, Ohio State University, 1957.

¹ Now at Central Michigan College.

² Sheila G. Morrison, "Measuring the Severity of Articulation Defectiveness," *JSHD*, XX (December, 1955), 347-351.

³ L. L. Thurstone and E. J. Chave, *The Measurement of Attitude* (Chicago, 1929).

⁴ D. Lewis and D. Sherman, "Measuring the Severity of Stuttering," *JSHD*, XVI (December, 1951), 320-326.

samples were presented a second time, the experimenter announcing the previously established level of severity of each in order that the judges might evaluate their practice judgments.

The foregoing procedure involving thirty-six samples was repeated and further practice was provided with the presentation of twenty additional segments.

A further similar training session of about forty minutes preceded each subsequent experimental judging session.

Apart from the training sessions the judges provided 4800 responses that were specific to the present study, one half relating to the pre-therapy samples and one half to the post-therapy samples. The mean of the pre-therapy ratings was 5.17 ± 1.58 and of the post-therapy ratings, 4.94 ± 1.71 . The difference between the two means was statistically significant at the 5 per cent level of confidence ($t, 2.05; 39$ d.f.).⁵ For the purpose of testing the reliability of judges, of items, or of subjects, either set of scores would serve equally well. Only the pre-therapy scores were used.

RELIABILITY OF THE JUDGES

Four procedures were followed in studying the reliability of the judges. Procedure 1 involved dividing the ten judges into odd-even groups and determining the Pearson product-moment correlation coefficient between the measures that represented the subgroups of judges. Similarly, eight of the judges, selected by chance from the ten, were subdivided and the correlation between the two sets of measures was computed. In like manner, six, four, and two were

⁵ A comparison of the two means is incidental to the present treatment. The product-moment correlation coefficient between the pre- and post-therapy scores for the forty subjects was .50.

selected and the correlation coefficient determined. By this procedure, the reliability coefficients of five, four, three, two, and one judge were found to be respectively: .95, .94, .90, .85, and .58, enumerated in the second column of Table I.

The foregoing correlations were made the bases of prediction through a series of applications of the Spearman-Brown prophecy formula. The predictions made by this formula hold if the test items are equivalent, i.e., if they have the same difficulty and if the item inter-correlations are equal,⁶ and may be used for stating the reliability of tests that are either longer or shorter than the original.⁷ Table I enumerates in columns 3-12 forty-five predicted reliability coefficients that were derived by applying the Spearman-Brown formula to reach possible N of judges to the observed values of column 2.

Inspection of Table I suggests that a sampling error occurred in the instance of one judge vs. one judge ($r = .58$). The forty-five possible comparisons of single judges are enumerated in Table II. Only seven of the values are as low as .58.

Thus in Procedure 1 an assumption was made that panels of 10, 8, 6, 4, and 2 judges had participated. In panels of fewer than ten members the "members" of the groups were selected randomly from the full panel of ten members. Split-half correlation was used in each instance to yield an observed indication of agreement between two parts of the sample. These five observed values, in turn, were made the bases for a series of predicted correlations for panels of

⁶ H. H. Remmers, "The Equivalence of Judgments to Test Items in the Sense of the Spearman-Brown Formula," *J. Ed. Psychol.*, XXII (January, 1931), 66-71.

⁷ Dorothy Adkins, *Construction and Analysis of Achievement Tests* (U. S. Govt. Printing Office, 1947).

TABLE I
PREDICTED VALUES FOR r FOR VARIOUS NUMBERS OF JUDGES USING OBSERVED PRODUCT-MOMENT COEFFICIENTS AS THE BASES FOR PREDICTION.

Number of Judges	Observed r					Predicted r				
	1	2	3	4	5	6	7	8	9	10
5	.95	.79	.88	.92	.94	—	.96	.96	.97	.97
4	.94	.80	.89	.92	—	.95	.96	.96	.97	.98
3	.90	.75	.86	—	.92	.94	.95	.95	.96	.97
2	.85	.74	—	.89	.92	.93	.94	.95	.96	.97
1	.58	—	.73	.81	.85	.87	.89	.91	.92	.93

TABLE II
AN INTERCORRELATION MATRIX OF THE RATINGS OF EACH JUDGE WITH EACH OF NINE OTHER JUDGES.

Judge Number	1	2	3	4	5	6	7	8	9	10
1	—	.81	.83	.80	.63	.67	.58	.74	.62	.61
2	—	—	.86	.82	.69	.72	.65	.72	.68	.49
3	—	—	.87	.74	.79	.75	.80	.72	.72	—
4	—	—	—	.72	.74	.73	.74	.69	.72	—
5	—	—	—	—	.52	.57	.58	.47	.53	—
6	—	—	—	—	—	.79	.97	.89	.86	—
7	—	—	—	—	—	—	.95	.85	.79	—
8	—	—	—	—	—	—	—	.83	.85	—
9	—	—	—	—	—	—	—	—	—	.84
10	—	—	—	—	—	—	—	—	—	—

judges ranging from one to ten members.

Apparently a panel of four or five judges of the level of the present sample would be sufficient to yield a reliability of $r = .92$ to $.95$, often a satisfactorily high value in rating situations. It should be remembered that these judges were trained by Morrison's method.

Procedure 2 involved one method of computing average intercorrelation. Since, for example, the value $.79$ in Table I was obtained by applying the prophecy formula to the coefficient obtained by a 5-5 division of the judging panel, it may be considered an estimate of the reliability of one random judge from the panel, or an estimated correlation between the ratings of any two randomly selected judges. This kind of value, more precise than the one provided by split-half correlation and termed *average intercorrelation*, was computed by means of a formula offered

by Edgerton and Toops.⁸ Historically, their development obviated the need for either working with ranked data (Kelly derived a procedure for determining average intercorrelation from ranked data⁹) or constructing an intercorrelation matrix and averaging the series. Computationally the Edgerton-Toops procedure is cumbersome. In the present instance it yielded an average intercorrelation coefficient of $.73$, the same value that might be obtained by averaging the five values of Table I that pertain to one judge: $.79$, $.80$, $.75$, $.74$, and $.58$. This value ($.73$), in turn, if extended by the Spearman-Brown formula would yield essentially the values of row 4 of Table I. Of especial relevance, however,

⁸ Harold A. Edgerton and Herbert A. Toops, "A Formula for Finding the Average Intercorrelation Coefficient of Unranked Raw Scores Without Solving Any of the Individual Intercorrelations," *J. Ed. Psychol.*, XIX (February, 1928), 131-138.

⁹ Truman L. Kelley, *Statistical Method* (New York, 1929).

the forty-five values of Table II illustrate ones that were superseded by the single formula for average intercorrelation, since they are coefficients of correlation of the ratings of each judge with those of each other judge. An average intercorrelation might be computed from them simply by averaging them. In the present instance the mean of these forty-five measures is .73, identical to the value yielded by the Edgerton-Toops formula.

Procedure 3 was the computation of the average intercorrelation coefficient by an analysis-of-variance technique, using a formula developed by Horst¹⁰ and rewritten and extended by Ebel and Hoyt:^{11, 12}

$$\bar{r} = \frac{(\text{Variance for subjects or items}) \text{ minus } (\text{remainder variance})}{(\text{Variance for subjects or items}) \text{ plus } (N-1 \text{ judges } [\text{remainder variance}])}$$

In an instance like the present one, the data for the analysis are entered on computation sheets in the same manner as for the foregoing Edgerton-Toops procedure: rows represent persons being judged and columns represent judges. The total variance is divided among columns (variance for judges), rows (variance for experimental subjects or items), and remainder. The procedure is an application of intraclass correlation.

In intraclass correlation, typically, the entries in a row of the matrix are merely samples from a common source. For example, the entries might be the number of words per page of a book. In this example different books would be represented in different rows. Columns would have little in common. Measures within a row, however, might be quite

similar and might be shifted to the right or left at will. The variance would be divided between *rows* and *remainder*.

The entries in the present matrix were the individual judges' pooled ratings for the six items that originated with one subject. This situation is hardly the one described in the preceding paragraph. Columns in the present case are expected to be stratified, each judge's ratings being more like each other than like his fellows' ratings. As disparity among judges (columns) increases within a particular matrix of values, the remainder or error variance decreases. For example, four hypothetically "similar" judges might lead to the following frac-

$$\text{tion: } \frac{100 - 25}{100 + 3(25)} = \frac{75}{175}$$

Four hypothetically dissimilar judges might cause the small values ("soft" judges) to fall in particular columns and the large numbers of the "hard" judges to fall in other columns. A larger proportion of the total variance would be attributed to *judges* and a smaller proportion to *remainder*. The fraction might become

$$\frac{100 - 15}{100 + 3(15)} = \frac{85}{145}$$

Thus, a high intraclass correlation value may suggest that a judge (column) is consistent with himself from judgment to judgment and that the values within a row may not be juxtaposed with a negligible effect upon the outcome. This characteristic of the rows of a matrix (equality of the values in a row) that is taken for granted in intraclass correlation is hardly expected in the rating circumstance. The apparent discrepancy, however, between the intraclass correlation and doubly stratified scores (strati-

¹⁰ Paul Horst, "A Generalized Expression for the Reliability of Measures," *Psychom.*, XIV (March, 1949), 21-31.

¹¹ Robert L. Ebel, "Estimation of the Reliability Ratings," *Psychom.*, XIV (December, 1951), 407-424.

¹² *Ibid.*

fied within columns and stratified within rows) is minimized with a rows \times columns treatment of the data so that Guilford speaks of intraclass correlation as "essentially an average intercorrelation."¹³

The method under discussion yielded

$$\bar{r} = \frac{(\text{Variance for subjects or items}) - (\text{remainder variance})}{\text{Variance for subjects or items}}$$

a correlation coefficient for a random member of the judging panel of .72, essentially the same as the other obtained values. Were this to be extended by the Spearman-Brown formula the results would be identical to ones already obtained. (This result may be checked by applying the formula to the data for ten judges from Table III.) The same formula was applied to successively smaller samples of the data as one judge (column) was dropped at random, two judges were dropped at random, etc. The obtained values were: from the data of nine judges, .71; from the data of eight judges, .72; from seven judges, .71; from six judges, .67; five judges, .73; four judges, .63; three judges, .68; and two judges, .71. The data from which these values were computed are enumerated in Table 3.

might be further employed in lieu of split-half correlation and the Spearman-Brown formula to estimate the reliability of N judges if the judgments are at hand. Ebel, following Hoyt, proposed the following formula for this purpose (the same numerator as above):

The estimated reliabilities for N = two randomly selected judges from the group of ten, three randomly selected judges, . . . and ten judges were: two judges, .83; three judges, .86; four judges, .87; five judges, .93; six judges, .92; seven judges, .94; eight judges, .95; nine judges, .96; and ten judges, .96. These are reliabilities for panels, not for an average member of the panel.

The foregoing material has treated three procedures for appraising the reliability of the judges who, after practicing with scaled samples, attempted to assign values to additional samples of children's articulation. The methods are (a) split-half product-moment correlation, (b) Edgerton-Toops average intercorrelation, and (c) analysis of variance. All of the estimates can be extended to larger and smaller samples by means of

TABLE III
SUMMARY OF THE DISTRIBUTION OF VARIANCE AMONG THE RATINGS OF 10 JUDGES AND OF
9, 8, 7, 6, 5, 4, 3, 2 JUDGES SELECTED RANDOMLY FROM THE 10.
(Degrees of freedom are indicated in parentheses)

Number of Judges		2	3	4	5	6	7	8	9	10
Source of Variation										
Judges = Columns										
.32 (2)	21.84 (2)	13.49 (3)	9.79 (4)	10.51 (5)	3.63 (6)	7.61 (7)	7.52 (8)	7.25 (9)		
Subjects = Rows (39)										
4.68	8.29	10.16	15.79	14.66	16.68	22.09	22.11	25.59		
Remainder										
.78 (39)	1.18 (78)	1.28 (117)	1.11 (156)	1.12 (195)	.94 (234)	1.01 (273)	.95 (312)	.95 (351)		

The procedure employed in computing the foregoing intraclass correlations and the obtained estimates of variance

the Spearman-Brown prophecy formula. Also, the use of analysis of variance for stating the reliability of an average judge leads to the application of a closely related technique, also analysis of variance,

¹³ J. P. Guilford, *Psychometric Methods* (sec. ed., New York, 1954).

for stating the reliability of a panel of judges.

Finally, the Spearman-Brown formula may be invoked to solve for N judges that would be expected to yield any desired coefficient of reliability, one of the obtained reliabilities for the average judge being inserted into the formula as a "known" value:

$$N = \frac{(\text{Desired coefficient}) (1 - \text{observed coefficient})}{(\text{Observed coefficient}) (1 - \text{desired coefficient})}$$

Thus, using the Edgerton-Toops value from above as the observed coefficient and arbitrarily selecting .90 as a satisfactory outcome, the solution for the data used here is:

$$N = \frac{.90 (1 - .73)}{.73 (1 - .90)} = 3.26 \text{ judges}^{13}$$

RELIABILITY OF TEST ITEMS

The items in the present test were ten-second samples of speech from which the speech of the source of the six items was to be evaluated. Obviously, this test should be no longer than it need be to measure the speaker reliably. The number of items or speech samples required for this score is a function of the reliability of the items. Two procedures were used to evaluate this.

First, the split-half procedure was applied, the pooled responses of ten judges to the *odd items* being correlated with the pooled ratings of the judges to the *even items* (r , .95). Two items, chosen randomly for each subject, were then discarded and the odd-even split-half technique applied again. Similarly, four items per subject were discarded. These observed correlations, .95, .93, and .83—shown in Table 4—were inserted into the Spearman-Brown formula and the reliability of one item, of two items, . . . of six items was predicted. The results are detailed in Table IV.

Second, the analysis-of-variance procedures described above in connection

with the reliability of judges were applied to six, five, four, three, two, and one item and the Spearman-Brown prophecy formula applied to each as a basis for prediction. The observed and predicted values are enumerated in Table V. Row 1 of Table V includes the reliability of one typical item (.85). This value is presumably more repre-

sentative of the matrix than is .83 of Row 1 of Table IV. The latter was derived through a procedure that included sampling; the former was computed from all of the data. This observation pertains also to the values associated with two items and three items in Tables IV and V.

The prediction formula that was used to determine the number of judges that would be required to maintain a reliability of .90 was applied to determine the number of items that would achieve this reliability and yielded 2.0 items. Table V indicates the consistency of different items and the extent to which a single ten-second item reveals a child's articulation.

RELIABILITY OF SUBJECTS

Attempts to predict by the Spearman-Brown prophecy formula the number of subjects required to yield specified results have not been successful.^{14,15} The split-half technique is hardly applicable in the present instance. The data may be envisaged as aligned in ten rows (judges) and two columns, twenty subjects pooled in each value. The outcome of this correlation would tell more about judges than about subjects.

¹⁴ Lyle H. Lanier, "Prediction of the Reliability of Mental Tests and Tests of Special Abilities," *J. Exp. Psychol.*, X (April, 1927), 69-113.

¹⁵ L. L. Thurstone, "A Note on the Spearman-Brown Formula," *J. Exp. Psychol.*, XI (February, 1928), 62-63.

TABLE IV
OBSERVED AND PREDICTED VALUES FOR r FOR VARIOUS NUMBERS OF ITEMS USING OBSERVED PRODUCT-MOMENT COEFFICIENTS AS THE BASES FOR PREDICTION BY THE SPEARMAN-BROWN FORMULA.

Numbers of Items	Observed r	Predicted r					
		1	2	3	4	5	6
1	.83	—	.91	.94	.95	.96	.97
2	.93	.87	—	.95	.96	.97	.98
3	.95	.86	.93	—	.96	.97	.97

TABLE V
OBSERVED AND PREDICTED VALUES FOR r FOR VARIOUS NUMBERS OF ITEMS. OBSERVED COEFFICIENTS COMPUTED BY ANALYSIS OF VARIANCE. PREDICTED VALUES COMPUTED BY SPEARMAN-BROWN FORMULA.

Number of Items	Observed r^*	Predicted r					
		1	2	3	4	5	6
1	.85	—	.92	.94	.96	.93	.97
2	.93	.87	—	.95	.96	.97	.98
3	.95	.84	.91	—	.96	.96	.96
4	.95	.84	.91	.94	—	.96	.96
5	.96	.85	.92	.95	.96	—	.97
6	.97	.85	.92	.95	.96	.96	—

*The observed value of r for $N = 1$ is an average intercorrelation (\bar{r}) computed by intraclass correlation; other observed values (r) are computed by the second analysis of variance formula cited above, the one with a single term in the denominator.

The analysis-of-variance procedure that was followed above in the treatment of judges and items offers some possibility for determining the reliability of subjects. The matrix suggested above would be spread out over forty columns (subjects). First the average inter-correlation (intraclass correlation) would be determined (.14). Predicted correlations from this value for various numbers of subjects are presented in the bottom row of Table VI. The matrix was then reduced by five columns (subjects) and the intraclass correlation recomputed. This process was continued for randomly selected populations of 5, 10, 15, 20, 25, 30, and 35 subjects from the entire matrix, and each of the obtained average intercorrelations (intraclass correlation) was used as a basis for predicting the reliability of groups of subjects of the six other sizes. These values, obtained and predicted, are presented in Table VI. The several vari-

ances that enter into the computation are presented in Table VII.

The observed r 's in Table VI are average intercorrelations computed by the Ebel analysis-of-variance technique for intraclass correlation. These values indicate estimates of the reliability of an average subject in a group of 5, 10, 15, 20, 25, 30, 35, as well as 40 subjects. The low average intercorrelation of the subjects is in keeping with an *individual-differences* and a *random-factor* concept of subjects. Each value, in turn, was used as the basis for a series of predictions of the reliability of populations of other sizes. An illuminating comparison arises from (a) these predicted measures and (b) the estimates of reliability yielded by the Ebel-Hoyt analysis-of-variance technique for stating the reliability of a panel. These obtained values were:

For five subjects	.45
For ten subjects	.52
For fifteen subjects	.78
For twenty subjects	.64

TABLE VI

OBSERVED AVERAGE INTERCORRELATION VALUES FOR AN AVERAGE SUBJECT OF FORTY SUBJECTS AND OF SEVEN GROUPS OF SUBJECTS DRAWN FROM THE FORTY AND THE PREDICTED RELIABILITY OF GROUPS OF SEVEN SIZES BASED ON EACH OBSERVED VALUE.

Number of Subjects	Observed \bar{r}				Predicted r			
	5	10	15	20	25	30	35	40
5	.08	—	.15	.21	.26	.31	.34	.38
10	.19	.10	—	.26	.32	.37	.41	.45
15	.19	.07	.13	—	.16	.29	.32	.35
20	.08	.02	.04	.06	—	.09	.11	.13
25	.13	.03	.06	.08	.11	—	.15	.19
30	.15	.03	.05	.08	.19	.17	—	.18
35	.14	.02	.05	.07	.08	.10	.12	—
40	.14	.02	.04	.06	.08	.09	.11	.12

TABLE VII

SUMMARY OF THE DISTRIBUTION OF VARIANCE AMONG THE RATINGS OF 10 JUDGES RELATIVE TO 40 SUBJECTS AND TO 5, 10, 15, 20, 25, 30, AND 35 SUBJECTS RANDOMLY SELECTED FROM THE 40.
(Degrees of freedom are indicated in parentheses)

Number of Subjects	5	10	15	20	25	30	35	40
Source of Variation								
Subj. =								
Cols.	24.06 (4)	27.22 (9)	21.62 (14)	28.51 (19)	25.01 (24)	23.24 (29)	24.40 (34)	25.59 (39)
Judges =								
Rows (9)	1.26	1.90	3.66	2.30	5.06	4.89	6.52	7.25
Remainder	.87 (36)	.91 (81)	.79 (126)	.83 (171)	1.04 (216)	.76 (261)	.97 (306)	.95 (351)

For twenty-five subjects

.92

For thirty subjects

.84

For thirty-five subjects

.85

For forty subjects

.97

A lack of systematic growth of the measures of reliability with larger numbers of subjects indicates the sensitivity of the procedure to sampling error.

DISCUSSION

An effect of this study is to highlight three sources of relative reliability in a rating procedure such as the one illustrated here, a rating of children's articulation. This reliability, in turn, may be computed and stated in several ways, equally appropriately.

An object in the present instance was to ferret out feasible economies in the testing procedure that was under study. With ten judges, the six items could have been reduced by two thirds, or with six items the judges could have been reduced by two thirds and a reliability of .90 maintained. Compound-

ing these sources of economy would presumably exceed the tolerable limits. An assumption was made that items and judges were equal in affecting the reliability of the test and that a reliability of .90 might be maintained by reducing both by one third (six judges and four items). Accordingly, the total matrix was reduced through random selection to these proportions and two split-half correlations were determined, one based on three *versus* three judges and the other on two *versus* two items. The former yielded a coefficient of .87 and the latter .91, indicating that no further economizing was in order if a reliability coefficient of .90 was to be claimed and tending to confirm that items and judges were approximately equally contributory to test reliability.

CONCLUSIONS

1. The kind of test material employed in appraising children's articulation yields judgmental data that fall within

the range of application of the Spearman-Brown prophecy formula in terms of test item and judge reliability, but less so with respect to subjects.

2. Equivalent statements of reliability may be obtained, in instances like the present one, from comparing parts (e.g. halves) of a test or from computing an average intercorrelation (or by averaging $N(N-1)/2$ individual correlations). The average intercorrelation can be computed by a Z transformation or by analysis of variance. Similarly there may be alternative procedures for stating the reliability of a population. This value may then be used to predict the reliability of a subsample of the population.

SUMMARY

Data from an articulation rating test were evaluated with respect to reliability. The procedures that were employed included (a) split-half and other seg-

mentings of the measures from ten judges, forty subjects, and six items per subject (the entire set of measures was also replicated); (b) average intercorrelation of the judges by means of a Z-score transformation of the data; and (c) average intercorrelation by means of analysis of variance (intraclass correlation). Interpretation of the obtained values was facilitated by an extensive use of the Spearman-Brown prophecy formula applied to measures obtained by the three foregoing procedures. These values, in turn, were in some instances compared with measures of reliability that were computed by analysis of variance and intended to state the reliability of a "group," not an average individual.

The Spearman-Brown formula, known to predict when items are equivalent, was applicable in extending measures obtained from judges and items, not subjects.

SPEECH MONOGRAPHS—INDEX TO VOLUME XXV

Authors and Articles

Abel, James W. [qu] Monosyllables. 29

Asher, John William. Intelligibility Tests: A Review of Their Standardization, Some Experiments, and a New Test. 14.

Auer, J. Jeffery. Doctoral Dissertations in Speech: Work in Progress, 1958. 77.

Ball, Joe M. The Relationship between the Ability to Speak Effectively and the Primary Mental Abilities, Verbal Comprehension and General Reasoning. 285

Black, John W. *See* March, Nancy.

Brockhaus, Herman H. The Wisconsin Sequential Sampling Audience Analyzer. 1.

Cleary, James W. A Bibliography of Rhetoric and Public Address for the year 1957-77.

Crowell, Laura. Word Changes Introduced *Ad Libitum* in five Speeches by Franklin Delano Roosevelt. 229.

Crowell, Laura. *See* Scheidel, Thomas M.

Dow, Clyde W. Abstract of Theses in the Field of Speech, XIII. 84.

Fotheringham, Wallace C. *See* Utterback, William E.

Hudson, Roy Fred. Rhetorical Invention in Colonial New England. 215.

Irwin, John V. *See* Brockhaus, Herman H.

Knower, Franklin H. Graduate Theses—An Index of Graduate Work in Speech, XXV. 151.

Lang, Robert A. Rhetoric in *Les Petites-Ecoles* of Port-Royal. 208.

Ludlum, Thomas S. Effects of Certain Techniques of Credibility Upon Audience Attitude. 278.

Mallory, Edith B. A Possible Basis for the Association of Voice Characteristic and Personality Traits. 255.

March, Nancy. Observed and Predicted Estimates of Reliability of Aspects of a Speech Articulation Rating Scale. 296.

McClendon, Paul I. An Experimental Study of the Relationship Between the Note-Taking Practices and Listening Comprehension of College Freshmen During Expository Lectures. 222.

McIntyre, Barbara M. The Effect of Creative Activities on the Articulation of Children with Speech Disorders. 42.

Miller, Virginia R. *See* Mallory, Edith B.

Morrison, Sheila G. *See* March, Nancy.

Nadeau, Ray. Hermogenes on "Stock Issues" in Deliberative Speaking. 59.

Nebergall, Roger E. An Experimental Investigation of Rhetorical Clarity. 243.

Phillips, Gerald M. Use of Authorities as Ethical Proof in the Talmudic Discourse. 67.

Ruechelle, Randall C. An Experimental Study of Audience Recognition of Emotional and Intellectual Appeals in Persuasion. 49.

Scheidel, Thomas M. Personality and Behavior: A Study of Possible Relationships. 261.

Shepherd, John R. *See* Scheidel, Thomas R.

Summers, Raymond. Oral and Nasal Sound Pressure Levels of Vowels. 291.

Utterback, William E. Experimental Studies of Motivated Group Discussion. 268.

Weaver, Carl H. *See* March, Nancy.

Authors of Abstracts

Abbott, Thomas B.	135
Aronson, Arnold E.	135
Austen, Albert Angus	98
Berquist, Goodwin F. Jr.	99
Bilsborrow, Eleanor Janice	100
Borriello, John Francis	84
Bowler, Ned Welcome	84
Boyer, Ernest L.	85
Brutten, Eugene Jerome	136
Burgi, Ernest J.	137
Byers, Burton H.	147
Campbell, Pearl Read	148
Cashman, Paul Harrison	88
Castleberry, Ottis L.	100
Costello, Mary Rose	137
Cox, Charles Wright	125
Dee, James Phillip	86
Draper, Walter Headen	126
Dreyfus, Lee Sherman	101
Drum, Dale D.	87
Dukore, Bernard Frank	126
Eubanks, Ralph T.	102
Evans, Patricia Morford	120
Fulsom, Ralph Edward	121
Foxen, John	102

Gee, Ronald C.	127
Gray, Gordon L.	122
Gregory, William Alfred, Jr.	128
Gregg, Robert LeRoy	122
Grissinger, James Adams	88
Grossman, Audley Mitchell, Jr.	128
Hardigree, Cruz Aurelia Cancel F.	88
Heller, Erma May Testai	138
Holm, James N.	103
Hoshiko, Michael S.	89
Isaacson, Pauline H.	104
Jackson, James Harvey	90
Janssen, Ann M.	129
Jeffrey, Robert C.	105
Johnson, E. Orville	106
Johnson, William Eddie	139
Kreps, Leslie Ray	106
Kyker, Rex Paxton	107
Larson, Gerard Arthur	129
Leistner, Charley A.	108
Ludlow, Jim Hand	123
McCurdy, Frances Lea	108
McIntyre, Barbara May	139
McLeod, Alan Lindsey	109
Miller, Charles James	130
Miller, Melvin Hull	110
Miller, Robert Warburton	90
Miller, Thomas Eugene	91
Monnier, Charles Martin	110
Munn, Mark D. II	124
Mysak, Edward D.	92
Nober, Earl Harris	92
Norman, Richard Arthur	93
North, Ross Stafford	111
Pace, Thomas Jennings, Jr.	112
Pennington, Paul Jordon	112
Phelps, Bernard Fred	113
Plotkin, William Harold	140
Preston, Lilliam Elvira	131
Quigley, Stephen Patrick	94
Ranney, J. Buckminster	95
Redding, William Charles	113
Reynolds, Wynn Robert	114
Rothman, Richard Marvin	115
Sabin, Cornelius Ayer	116
Saito, Mitsuko	148
Scholl, Harold M.	141
Schwartz, Ralph Jerome	142
Shanks, James Clements, Jr.	95
Shea, William L.	142
Sklar, Maurice	143
Smith, Robert Wayne	116
Sonday, Francis Llewellyn	144
Stallings, Juadina Brock	117
Stelzner, Hermann Georg	118
Stephenson, Jim Bob	132
Summers, Dorothy Iris	118
Teegarden, James Carr	96
Thomas, Wanda Clayton	132
Thompson, Jesse J.	145
Tikofsky, Ronald Sherwood	145
Torrence, Donald Lee	119
Turner, Daniel	97
Veilleux, Jere Shanor	133
Wattron, Frank Joseph, Jr.	133
Weaver, Carl H.	97
Webb, Clarence Elmer	146
White, Charles Arthur	120
Wilson, Mardis Glen	134
Winchester, Richard A.	146
Woodlley, Thomas Russel, Jr.	124
Young, Norton B.	98

.. 113
.. 140
.. 131
.. 94
.. 95
.. 113
.. 114
.. 115
.. 116
.. 148
.. 141
.. 142
.. 95
.. 142
.. 143
.. 116
.. 144
.. 117
.. 118
.. 132
.. 118

.. 96
.. 132
.. 145
.. 145
.. 119
.. 97
.. 133

.. 133
.. 97
.. 146
.. 120
.. 134
.. 146
.. 124

.. 98